

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2003, 22:55:32 : Search time 266 seconds
(without alignments)
194.722 Million cell updates/sec

Title: SEQ1-T-AT-12347_COPY_12336_12358

Perfect score: 23

Sequence: 1 ggggcaaatctactctgcctcc 23

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 112599159 residues

Word size : 12

Total number of hits satisfying chosen parameters: 1343

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002:*

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	16	69.6	700	AAH92847	Human inflammatory
C 2	15	65.2	420	AAH74522	Human CDNA clone #
C 3	15	65.2	462	AA57386	CDNA #62 encoding
C 4	15	65.2	588	AAK60492	Human immune/haema
C 5	15	65.2	630	AAH07927	Human mtchochondria
C 6	15	65.2	946	AAH5428	Nucleotide sequenc
C 7	15	65.2	1005	AAH57485	DNA encoding novel
C 8	15	65.2	99960	AAZ50905	Human TBC-1 partia
C 9	14	60.9	236	ABA12200	Human nervous syst

10	14	60.9	660	24	ABK76032	Bacillus lichenito
C 11	14	60.9	753	24	ABQ47496	oligonucleotide fo
C 12	14	60.9	753	24	ABQ47497	oligonucleotide fo
C 13	14	60.9	767	24	AHQ23634	oligonucleotide fo
C 14	14	60.9	767	24	AHQ23635	oligonucleotide fo
C 15	14	60.9	791	23	AAH86152	DNA encoding novel
C 16	14	60.9	966	24	ABL87945	DNA polymerase III
C 17	14	60.9	966	24	ABL87946	DNA polymerase III
C 18	14	60.9	2622	24	ABK63737	Kel sequence diffe
C 19	14	60.9	2637	22	AA160113	Human polynucleoti
C 20	14	60.9	3610	23	ABL27970	Human polynucleoti
C 21	14	60.9	4780	24	ABA91370	Human breast speci
C 22	14	60.9	6070	24	ABL92198	Chemically treated
C 23	14	60.9	6070	24	ABL49309	Human polynucleoti
C 24	14	60.9	6070	24	ABL32240	Human immune syste
C 25	14	60.9	6120	23	ABL29702	Human immune syste
C 26	14	60.9	6300	23	ABL29704	Human immune syste
C 27	14	60.9	7195	22	AAH45324	Human immune syste
C 28	14	60.9	7195	22	AAH28165	Human immune syste
C 29	14	60.9	7353	24	ABL32073	Human immune syste
C 30	14	60.9	7353	24	ABD28363	Human chemically t
C 31	14	60.9	8047	24	ABH80325	Human chemically t
C 32	14	60.9	9762	23	ABL04410	Human chemically t
C 33	14	60.9	17908	22	AAH37254	Human chemically t
C 34	14	60.9	17908	22	AAH37254	Human chemically t
C 35	14	60.9	35099	19	AAV27112	Human chemically t
C 36	14	60.9	40267	24	AAH18958	Human chemically t
C 37	14	60.9	100301	24	ABO88176	Human osteoblast d
C 38	14	60.9	3011208	24	ABO69245	Listeria innocua D
C 39	13	56.5	27	20	AAH37795	Staphylococcus sp.
C 40	13	56.5	97	16	AAH24854	Human gene signatu
C 41	13	56.5	100	16	AAQ97384	Human type I stero
C 42	13	56.5	163	16	AAH20244	Human gene signatu
C 43	13	56.5	197	21	AAH20244	Human gene signatu
C 44	13	56.5	209	21	AAH20244	Human gene signatu
C 45	13	56.5	211	20	AAH40460	Human secreted pro

ALIGNMENTS

RESULT 1	
AAH92847/c	
AAH92847 standard; DNA: 700 bp.	
XX	
XX	
AC	AAH92847:
XX	
DT	09-OCT-2001 (first entry)
XX	
XX	
DE	Human inflammatory bowel disease related gene fragment (c1172a).
XX	
KW	Human: int inflammatory bowel disease; Crohn's disease; ulcerative colitis;
KW	Single nucleotide polymorphism; SNP; chromosome 19p13; paternally test;
KW	chromosome 5q31-33; forensic test; gene therapy; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200142511-A2.
XX	
PD	14-JUN-2001.
XX	
PF	11-DEC-2000; 2000WO-US33632.
XX	
PR	10-DEC-1999; 9905-0170257.
PR	10-APR-2000; 2000US-0196046.
XX	
PA	(WHED) WHITEHEAD INST BIOMEDICAL RES.
PA	(ELLI-) ELLIPSIS BIOTHERAPEUTICS CORP.
XX	
PI	Daily M, Hudson TJ, Lander ES, Rioux J, Stiminovich K;
XX	
DR	WPL; 2001-367874/38.

PT Testing for the presence of polymorphisms associated with inflammatory
 PT bowel disease, using a hybridization assay -
 XX
 PS Disclosure: Page 331-332; 463pp: English.

CC The present invention describes a method for detecting the presence of
 CC polymorphisms associated with inflammatory bowel diseases such as
 CC ulcerative colitis and Crohn's disease. The methods can be used to detect
 CC the presence of genetic polymorphisms associated with inflammatory bowel
 CC disease and correlating their occurrence with disease states. They may be
 CC used in this way for phenotypic correlations, forensics, paternity
 CC testing, medicine and genetic analysis. The present sequence is a gene
 CC containing a polymorphic site described in the exemplification of the
 CC invention.

XX Sequence 700 BP: 274 A; 88 C; 115 G; 219 T; 4 other;

Query Match 69.6%; Score 16; DB 22; Length 700;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGCGCAATCTTACTTT 17
 |||||||
 DB 385 GGCGCAATCTTACTTT 370

RESULT 2

AAA74522/c
 ID AAA74522 standard; cDNA: 420 BP.

AC AAA74522;

DT 05-DEC-2000 (first entry)

XX Human cDNA clone #1 used for designing probes.

XX Human; probe design: low cross hybridisation; chemical;

KW biological synthesis; diagnostic; therapeutic; ss.

XX Homo sapiens.

OS WO200043942-A2.

XX 27-JUL-2000.

XX 25-JAN-2000; 2000WO-US02000.

XX 25-JAN-1999; 99US-0116956.

XX (COMB-) COMBIMATRIX CORP.

XX Anderson RP, Montgomery DU;

XX WPI: 2000-491203/43.

XX Selecting oligonucleotide probes with low cross-hybridization, useful
 PT for therapy or diagnosis, by testing candidate probe for
 PT cross-reactivity with non-target sequences -

XX Example 1: Page 9; 18pp: English.

CC The present sequence is a human cDNA clone. This sequence was used for
 CC designing probes for use in low cross hybridisation i.e. the probes
 CC hybridise to their intended targets, but not to other targets. The
 CC resulting probes are useful in chemical or biological synthesis,
 CC diagnostics and therapeutics.

XX Sequence 420 BP: 102 A; 108 C; 83 G; 127 T; 0 other;

Query Match 65.2%; Score 15; DB 21; Length 420;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGCGCAATCTTACTT 16
 |||||||
 DB 340 GGCGCAATCTTACTT 326

RESULT 3

AAS57386/c
 ID AAS57386 standard; cDNA: 462 BP.

AC AAS57386;

DT 13-FEB-2002 (first entry)

XX cDNA #62 encoding portion of a human colon tumour protein.

XX Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.

XX Homo sapiens.

XX WO200173027-A2.

XX 04-OCT-2001.

XX 22-MAR-2001; 2001WO-US09246.

XX 24-MAR-2000; 2000US-191597P.

XX 04-MAY-2000; 2000US-202024P.

XX 05-MAY-2000; 2000US-202189P.

XX (CORI-) CORIXA CORP.

XX Meagher MJ, Xu J, King GE;

XX WPI: 2001-611627/70.

XX Claim 4; Page 71; 299pp: English.

CC Th present invention relates to the isolation of novel cDNA sequences
 CC encoding for at least an immunogenic portion of human colon tumour
 CC proteins. The sequences of the invention are useful in pharmaceutical
 CC compositions and vaccines for the prevention and treatment of cancers
 CC such as colon cancer. They are also useful for the diagnosis and
 CC monitoring of such cancers. Antibodies to the colon tumour proteins
 CC and antigen presenting cells that express polynucleotides encoding
 CC colon tumour proteins can be used to inhibit the development of
 CC cancers. T-cells that react specifically with colon tumour proteins
 CC are useful for removing tumour cells from samples (e.g. blood) and
 CC for cancer treatment. The polynucleotides sequences are also useful in
 CC gene therapy. AAS57325-AAS58880 represent the cDNA sequences of the
 CC invention that encode for portions of human colon tumour proteins.

XX Sequence 462 BP: 114 A; 115 C; 92 G; 132 T; 9 other;

Query Match 65.2%; Score 15; DB 23; Length 462;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGCGCAATCTTACTT 16
 |||||||
 DB 313 GGCGCAATCTTACTT 299

RESULT 4

AAK60492/c
 ID AAK60492 standard; cDNA: 588 BP.

AC AAK60492;

XX 06-NOV-2001 (first entry)

DE: Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:5552.
XX
XX Human: immune: haematopoietic: immune/haematopoietic antigen: cancer:
KW cytostatic: gene therapy: vaccine: metastasis: ss.
XX
XX Homo sapiens.
XX
XX W0200157182-A2.
XX
PD 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218296.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0227558.
PR 14-AUG-2000; 2000US-0227559.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227099.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229511.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0231415.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 27-SEP-2000; 2000US-0235837.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239397.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249300.
PR 17-NOV-2000; 2000US-0249309.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.

08-DEC-2000: 2000US-0251990.
 PR 11-DEC-2000: 2000US-0254097.
 PR 05-JAN-2001: 2001US-0259678.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI: 2001-483426/52.
 DR P-PSDB: AAM87711.
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX
 PS Claim 1; SEQ ID NO 5552: 3071bp + Sequence listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (1)
 CC amino acid sequences given in AAM82170 to AAM91921. (1) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (1)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (1) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (1) by expressing inactive proteins or to
 CC supplement the patient's own production of (1). Additionally, (1)
 CC polynucleotides may be used to produce the secreted (1), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (1) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 588 BP; 183 A; 114 C; 128 G; 160 T; 3 other:
 Query Match 65.28; Score 15; DB 22; Length 588;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 3 GGCAATCTTACTTT 17
 ||||||||||||
 DB 86 GGCAATCTTACTTT 72
 ||||||||||||
 RESULT 5
 AAD07927/c
 ID AAD07927 standard; CDNA: 630 BP.
 XX
 AC AAD07927;
 XX
 DT 04-AUG-2001 (first entry)
 XX
 DE Human mitochondrial deformylase partial cDNA #2.
 XX
 KW Human: mitochondrial deformylase; neoplastic disease; cancer; AIDS;
 KW Acquired Immune Deficiency Syndrome; cell proliferation; cytosatic;
 KW neurotropic; tranquiliser; antisense gene therapy; muscular dystrophy;
 KW amyotrophic lateral sclerosis; autoimmune disease; leukaemia; mood;
 KW nervous system disorder; anxiety; neurodegenerative disorder;
 KW Alzheimer's disease; Parkinson's disease; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS complement (1..630)
 FT /tag=a
 FT /product="Human mitochondrial deformylase protein
 FT fragment #2"
 FT /transl_except= (pos:595..597, aa:Xaa)
 FT /transl_except= (pos:571..573, aa:Xaa)
 FT /transl_except= (pos:400..402, aa:Xaa)

/transl_except- (pos:262..264, aa:Xaa)
 FT /transl_except- (pos:229..231, aa:Xaa)
 FT /transl_except- (pos:109..111, aa:Xaa)
 FT /transl_except- (pos:52..54, aa:Xaa)
 FT /transl_except- (pos:37..39, aa:Xaa)
 FT /transl_except- (pos:4..6, aa:Xaa)
 FT /note="Xaa corresponds to in-frame stop codon: (1)S
 FT does not include start and stop codon"
 FT /partial
 XX
 PN W0200142431-A2.
 XX
 PD 14-JUN-2001.
 XX
 PF 01-DEC-2000: 2000MO-EP12110.
 XX
 PR 08-JUN-1999: 99US-0169615.
 PR 11-OCT-2000: 2000US-0239106.
 XX
 PA (FARB) BAYER AG.
 PI Ramakrishnan S;
 XX
 DR WPI: 2001-381664/40.
 DR P-PSDB: AAE03547.
 XX
 PT Treating neoplastic disease such as cancer by administering a reagent
 PT which modulates human mitochondrial deformylase activity or expression
 PT to modulate cell proliferation -
 XX
 PS Claim 10; Page 70: 73pp; English.
 XX
 CC The present sequence is a cDNA encoding human mitochondrial deformylase
 CC protein fragment. Mitochondrial deformylase is an enzyme which cleaves
 CC the formyl group from nascent formyl-methionine peptides. Mitochondrial
 CC deformylase is useful for treating neoplastic disease, AIDS, muscular
 CC dystrophy, amyotrophic lateral sclerosis, muscle wasting diseases and
 CC autoimmune diseases. Mitochondrial deformylase is useful for decreasing
 CC cell proliferation and is useful for treating cancers such as leukaemia,
 CC adenocarcinoma and other diseases involving increased levels of cell
 CC proliferation. Since mitochondrial deformylase is of importance to
 CC central and peripheral nervous system it is also useful as a target for
 CC treating nervous system disorders such as disorders of mood, anxiety
 CC disorders, disorders of thought and volition, disorders of sleep and
 CC wakefulness, neurodegenerative disorders such as Alzheimer's and
 CC Parkinson's disease. Mitochondrial deformylase DNA is also useful in
 CC antisense gene therapy.
 XX
 SQ Sequence 630 BP; 141 A; 176 C; 139 G; 171 T; 3 other:
 Query Match 65.28; Score 15; DB 22; Length 630;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 2 GGCCAAATCTTACTT 16
 ||||||||||||
 DB 338 GGCCAAATCTTACTT 324
 ||||||||||||
 RESULT 6
 AAF55428
 ID AAF55428 standard; DNA: 946 BP.
 XX
 AC AAF55428;
 XX
 DT 29-MAY-2001 (first entry)
 XX
 DE Nucleotide sequence of a human hydrolytic enzyme HYEN24.
 XX
 KW Human: hydrolytic enzyme; HYEN2; neurological disorder; cancer;
 KW immune system disorder; genetic disorder; cell proliferation disorder;
 KW epilepsy; ischemic cerebrovascular disease; stroke; Pick's disease;
 KW Huntington's disease; dementia; Parkinson's disease; multiple sclerosis;

KW viral meningitis; Creutzfeldt-Jakob disease; neurofibromatosis;
 KW cerebral palsy; autonomic nervous system disorder; mental disorder;
 KW cranial nerve disorder; peripheral nervous system disorder;
 KW immune system disorder; osteoarthritis; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 35..361
 FT /tag= a
 FT /product= "hydrolytic enzyme"
 XX
 PN WO200116334-A2.
 XX
 PD 08-MAR-2001.
 XX
 PF 31-AUG-2000; 2000WO-US24107.
 XX
 PR 01-SEP-1999; 99US-0151819.
 XX
 PA (INCYTE) INCYTE GENOMICS INC.
 XX
 PI Yue H, Hillman JL, Tang YT, Baughn MR, Lu DAM, Azimzai Y;
 DR WPI: 2001-235111/24.
 DR P-PSDB: AAB67572.
 XX
 PT Novel human hydrolytic enzymes useful for diagnosing, treating, or
 PT preventing disorders associated with abnormal expression of HYENZ, cell
 PT proliferative disorders, neurological disorders and immune system
 PT disorders -
 XX
 PS Claim 5: Page 106; 113pp; English.
 XX
 CC The present sequence encodes a human hydrolytic enzyme (HYENZ). The
 CC specification describes HYENZ-1 to HYENZ-14. HYENZ polypeptides and
 CC polynucleotides are useful in the diagnosis, prevention and treatment of
 CC neurological disorders, immune system disorders, genetic disorders, and
 CC cell proliferation disorders including cancer. They are useful for
 CC treating epilepsy, ischemic cerebrovascular disease, stroke, Pick's
 CC disease, Huntington's disease, dementia, Parkinson's disease, multiple
 CC sclerosis, viral meningitis, Creutzfeldt-Jakob disease,
 CC neurofibromatosis, cerebral palsy, autonomic nervous system disorder,
 CC cranial nerve disorder, peripheral nervous system disorder, mental
 CC disorders, immune system disorders, osteoarthritis, and genetic
 CC disorders. HYENZ polynucleotides are useful for somatic or germline
 CC gene therapy for treating the disorders.
 XX
 SO Sequence 946 BP; 253 A; 227 C; 267 G; 199 T; 0 other:
 Query Match 65.2%; Score 15; DB 22; Length 946;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GGGCAATCTTACTT 16
 DB 423 GGGCAATCTTACTT 437
 XX
 RESULT 7
 AAS75485
 ID AAS75485 standard; cDNA; 1005 BP.
 XX
 AC AAS75485;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #11289.
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.

XX
 PN WO2001175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 DR WPI: 2001-639362/73.
 DR P-PSDB: ABG11298.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1: SEQ ID NO 11289; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging or sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/publ/published_pat_sequences.
 XX
 SO Sequence 1005 BP; 176 A; 291 C; 355 G; 183 T; 0 other:
 Query Match 65.2%; Score 15; DB 23; Length 1005;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GGGCAATCTTACTT 16
 DB 794 GGGCAATCTTACTT 808
 XX
 RESULT 8
 AA250905/c
 ID AA250905 standard; DNA; 9960 BP.
 XX
 AC AA250905;
 XX
 DT 31-MAY-2000 (first entry)
 XX
 DE Human TBC-1 partial genomic DNA comprising 3' end sequence.
 KW TBC-1; human; diallelic marker; chromosome 4; cell cycle regulator; SNP;
 KW single nucleotide polymorphism; tissue differentiation; prostate cancer;
 KW linkage analysis; genetic map; detection; diagnosis; genotyping;
 KW transgenic animal; screening; ds.
 XX
 OS Homo sapiens.

XX	Key	Location/Qualifiers
FT	exon	4661..4789
FT		/*tag= a
FT	intron	/label= A
FT		4790..6115
FT		/*tag= b
FT	exon	/label= A
FT		6116..6202
FT		/*tag= c
FT	intron	/label= B
FT		6203..9918
FT		/*tag= d
FT	exon	/label= B
FT		9919..10199
FT		/*tag= e
FT	intron	/label= C
FT		10200..14520
FT		/*tag= f
FT	exon	/label= C
FT		14521..14660
FT		/*tag= g
FT	intron	/label= D
FT		14661..50256
FT		/*tag= h
FT	exon	/label= D
FT		50257..50442
FT		/*tag= i
FT	intron	/label= E
FT		50443..56255
FT		/*tag= j
FT	exon	/label= E
FT		56256..56417
FT		/*tag= k
FT	intron	/label= F
FT		56418..63325
FT		/*tag= l
FT	exon	/label= F
FT		63326..63484
FT		/*tag= m
FT	intron	/label= G
FT		63485..76035
FT		/*tag= n
FT	exon	/label= G
FT		76036..76280
FT		/*tag= o
FT	intron	/label= H
FT		76281..78363
FT		/*tag= p
FT	exon	/label= H
FT		78364..78523
FT		/*tag= s
FT	intron	/label= I
FT		78524..85294
FT		/*tag= t
FT	exon	/label= I
FT		85295..85464
FT		/*tag= u
FT	intron	/label= J
FT		85465..93416
FT		/*tag= v
FT	exon	/label= J
FT		93417..93590
FT		/*tag= w
FT	intron	/label= K
FT		93591..97475
FT		/*tag= x
FT	exon	/label= K
FT		97476..97960
FT		/*tag= y
FT	primer_bind	/label= L
FT		988..1006
FT		/*tag= z

FT		
FT	primer_bind	/bound_motety= "Primer B2"
FT		/note= "Amplification of amplicon 99-20508"
FT		complement (1509..1529)
FT		/*tag= aa
FT	primer_bind	/bound_motety= "Primer C2"
FT		/note= "Amplification of amplicon 99-20508"
FT		5039..5056
FT		/*tag= ab
FT	primer_bind	/bound_motety= "Primer H3"
FT		/note= "Amplification of amplicon 99-20469"
FT		complement (5534..5554)
FT		/*tag= ac
FT	primer_bind	/bound_motety= "Primer C3"
FT		/note= "Amplification of amplicon 99-20469"
FT		5997..6015
FT		/*tag= ad
FT	primer_bind	/bound_motety= "Primer B4"
FT		/note= "Amplification of amplicon 5-254"
FT		complement (6332..6350)
FT		/*tag= ae
FT	primer_bind	/bound_motety= "Primer C4"
FT		/note= "Amplification of amplicon 5-254"
FT		14371..14390
FT		/*tag= ai
FT	primer_bind	/bound_motety= "Primer B5"
FT		/note= "Amplification of amplicon 5-257"
FT		complement (14798..14817)
FT		/*tag= ag
FT	primer_bind	/bound_motety= "Primer C5"
FT		/note= "Amplification of amplicon 5-257"
FT		18751..18771
FT		/*tag= ah
FT	primer_bind	/bound_motety= "Primer B6"
FT		/note= "Amplification of amplicon 99-20511"
FT		complement (19198..19217)
FT		/*tag= ai
FT	primer_bind	/bound_motety= "Primer C6"
FT		/note= "Amplification of amplicon 99-20511"
FT		19605..19625
FT		/*tag= aj
FT	primer_bind	/bound_motety= "Primer B7"
FT		/note= "Amplification of amplicon 99-20510"
FT		complement (19986..20005)
FT		/*tag= ak
FT	primer_bind	/bound_motety= "Primer C7"
FT		/note= "Amplification of amplicon 99-20510"
FT		29529..29547
FT		/*tag= al
FT	primer_bind	/bound_motety= "Primer B8"
FT		/note= "Amplification of amplicon 99-20504"
FT		complement (40041..40061)
FT		/*tag= am
FT	primer_bind	/bound_motety= "Primer C8"
FT		/note= "Amplification of amplicon 99-20504"
FT		42268..42287
FT		/*tag= an
FT	primer_bind	/bound_motety= "Primer B9"
FT		/note= "Amplification of amplicon 99-20493"
FT		complement (42712..42752)
FT		/*tag= ao
FT	primer_bind	/bound_motety= "Primer C9"
FT		/note= "Amplification of amplicon 99-20493"
FT		69026..69046
FT		/*tag= ap
FT	primer_bind	/bound_motety= "Primer B10"
FT		/note= "Amplification of amplicon 99-20499"
FT		complement (69525..69543)
FT		/*tag= ag
FT	primer_bind	/bound_motety= "Primer C10"
FT		/note= "Amplification of amplicon 99-20499"
FT		76323..76343
FT		/*tag= ar
FT	primer_bind	/bound_motety= "Primer B11"

```
FT primer_bind /note= "Amplification of amplicon 99-20473"  
FT complement (76771..76790)  
FT /tag= as  
FT /bound_moiety= "Primer C11"  
FT /note= "Amplification of amplicon 99-20473"  
FT 78292..78309  
FT /tag= at  
FT /bound_moiety= "Primer B12"  
FT /note= "Amplification of amplicon 5-249"  
FT complement (78704..78721)  
FT /tag= au  
FT /bound_moiety= "Primer C12"  
FT /note= "Amplification of amplicon 5-249"  
FT 81893..81912  
FT /tag= av  
FT /bound_moiety= "Primer B13"  
FT /note= "Amplification of amplicon 99-20485"  
FT complement (82353..82372)  
FT /tag= aw  
FT /bound_moiety= "Primer C13"  
FT /note= "Amplification of amplicon 99-20485"  
FT 84392..84412  
FT /tag= ax  
FT /bound_moiety= "Primer B14"  
FT /note= "Amplification of amplicon 99-20481"  
FT complement (84909..84929)  
FT /tag= ay  
FT /bound_moiety= "Primer C14"  
FT /note= "Amplification of amplicon 99-20481"  
FT 89746..89765  
FT /tag= az  
FT /bound_moiety= "Primer B15"  
FT /note= "Amplification of amplicon 99-20480"  
FT complement (90179..90198)  
FT /tag= ba  
FT /bound_moiety= "Primer C15"  
FT /note= "Amplification of amplicon 99-20480"  
FT 1431..1455  
FT /tag= bb  
FT /bound_moiety= "Probe P2"  
  
Query Match 65.2%; Score 15; DB 21; Length 99960;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 5 CAATCTTACTTCG 19  
Db 10209 CAATCTTACTTCG 10195  
  
RESULT 9  
AA12200/c  
ID ABA12200 standard; cDNA: 236 bp.  
XX  
AC ABA12200;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Human nervous system related polynucleotide SEQ ID NO 1207.  
XX  
KW Human; neurotropic; antiinflammatory; cytosolic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
KW antiparkinsonian; antistickling; antianaemic; antiarthritic; cancer;  
KW antithematic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200159063-A2.  
XX  
XX 16-AUG-2001.  
PD
```

```
XX  
PF 17-JAN-2001: 2001WO-US01334.  
XX  
PR 31-JAN-2000: 2000US-0179065.  
PR 04-FEB-2000: 2000US-0180628.  
PR 24-FEB-2000: 2000US-0184664.  
PR 02-MAR-2000: 2000US-0186350.  
PR 16-MAR-2000: 2000US-0189874.  
PR 17-MAR-2000: 2000US-0190076.  
PR 18-APR-2000: 2000US-0198123.  
PR 19-MAY-2000: 2000US-0205515.  
PR 07-JUN-2000: 2000US-0209467.  
PR 28-JUN-2000: 2000US-0214886.  
PR 30-JUN-2000: 2000US-0215135.  
PR 07-JUL-2000: 2000US-0216647.  
PR 07-JUL-2000: 2000US-0216880.  
PR 11-JUL-2000: 2000US-0217487.  
PR 11-JUL-2000: 2000US-0217496.  
PR 14-JUL-2000: 2000US-0218290.  
PR 26-JUL-2000: 2000US-0220963.  
PR 26-JUL-2000: 2000US-0220964.  
PR 14-AUG-2000: 2000US-0224518.  
PR 14-AUG-2000: 2000US-0224519.  
PR 14-AUG-2000: 2000US-0225213.  
PR 14-AUG-2000: 2000US-0225214.  
PR 14-AUG-2000: 2000US-0225266.  
PR 14-AUG-2000: 2000US-0225267.  
PR 14-AUG-2000: 2000US-0225268.  
PR 14-AUG-2000: 2000US-0225270.  
PR 14-AUG-2000: 2000US-0225447.  
PR 14-AUG-2000: 2000US-0225757.  
PR 14-AUG-2000: 2000US-0225758.  
PR 14-AUG-2000: 2000US-0225759.  
PR 18-AUG-2000: 2000US-0226279.  
PR 22-AUG-2000: 2000US-0226681.  
PR 22-AUG-2000: 2000US-0226868.  
PR 22-AUG-2000: 2000US-0227182.  
PR 23-AUG-2000: 2000US-0227009.  
PR 30-AUG-2000: 2000US-0228924.  
PR 01-SEP-2000: 2000US-0229287.  
PR 01-SEP-2000: 2000US-0229343.  
PR 01-SEP-2000: 2000US-0229344.  
PR 01-SEP-2000: 2000US-0229345.  
PR 05-SEP-2000: 2000US-0229509.  
PR 05-SEP-2000: 2000US-0229513.  
PR 06-SEP-2000: 2000US-0230437.  
PR 08-SEP-2000: 2000US-0231242.  
PR 08-SEP-2000: 2000US-0231243.  
PR 08-SEP-2000: 2000US-0231244.  
PR 08-SEP-2000: 2000US-0231413.  
PR 08-SEP-2000: 2000US-0231414.  
PR 08-SEP-2000: 2000US-0232080.  
PR 08-SEP-2000: 2000US-0232081.  
PR 12-SEP-2000: 2000US-0231968.  
PR 14-SEP-2000: 2000US-0232397.  
PR 14-SEP-2000: 2000US-0232398.  
PR 14-SEP-2000: 2000US-0232399.  
PR 14-SEP-2000: 2000US-0232400.  
PR 14-SEP-2000: 2000US-0232401.  
PR 14-SEP-2000: 2000US-0233063.  
PR 14-SEP-2000: 2000US-0233064.  
PR 14-SEP-2000: 2000US-0233065.  
PR 14-SEP-2000: 2000US-0233066.  
PR 21-SEP-2000: 2000US-0234223.  
PR 21-SEP-2000: 2000US-0234274.  
PR 25-SEP-2000: 2000US-0234997.  
PR 25-SEP-2000: 2000US-0234998.  
PR 26-SEP-2000: 2000US-0235484.  
PR 27-SEP-2000: 2000US-0235834.  
PR 27-SEP-2000: 2000US-0235836.  
PR 29-SEP-2000: 2000US-0236327.  
PR 29-SEP-2000: 2000US-0236367.  
PR 29-SEP-2000: 2000US-0236368.
```

PR 29-SEP-2000: 2000US-0236369.
 PR 29-SEP-2000: 2000US-0236370.
 PR 02-OCT-2000: 2000US-0236802.
 PR 02-OCT-2000: 2000US-0237037.
 PR 02-OCT-2000: 2000US-0237038.
 PR 02-OCT-2000: 2000US-0237039.
 PR 02-OCT-2000: 2000US-0237040.
 PR 13-OCT-2000: 2000US-0239975.
 PR 13-OCT-2000: 2000US-0239977.
 PR 20-OCT-2000: 2000US-0240960.
 PR 20-OCT-2000: 2000US-0241785.
 PR 20-OCT-2000: 2000US-0241786.
 PR 20-OCT-2000: 2000US-0241787.
 PR 20-OCT-2000: 2000US-0241808.
 PR 20-OCT-2000: 2000US-0241809.
 PR 20-OCT-2000: 2000US-0241826.
 PR 20-OCT-2000: 2000US-0242221.
 PR 01-NOV-2000: 2000US-0244617.
 PR 08-NOV-2000: 2000US-0246474.
 PR 08-NOV-2000: 2000US-0246475.
 PR 08-NOV-2000: 2000US-0246476.
 PR 08-NOV-2000: 2000US-0246477.
 PR 08-NOV-2000: 2000US-0246478.
 PR 08-NOV-2000: 2000US-0246523.
 PR 08-NOV-2000: 2000US-0246524.
 PR 08-NOV-2000: 2000US-0246525.
 PR 08-NOV-2000: 2000US-0246526.
 PR 08-NOV-2000: 2000US-0246527.
 PR 08-NOV-2000: 2000US-0246528.
 PR 08-NOV-2000: 2000US-0246532.
 PR 08-NOV-2000: 2000US-0246609.
 PR 08-NOV-2000: 2000US-0246610.
 PR 08-NOV-2000: 2000US-0246611.
 PR 08-NOV-2000: 2000US-0246613.
 PR 17-NOV-2000: 2000US-0249207.
 PR 17-NOV-2000: 2000US-0249208.
 PR 17-NOV-2000: 2000US-0249209.
 PR 17-NOV-2000: 2000US-0249210.
 PR 17-NOV-2000: 2000US-0249211.
 PR 17-NOV-2000: 2000US-0249212.
 PR 17-NOV-2000: 2000US-0249213.
 PR 17-NOV-2000: 2000US-0249214.
 PR 17-NOV-2000: 2000US-0249215.
 PR 17-NOV-2000: 2000US-0249216.
 PR 17-NOV-2000: 2000US-0249217.
 PR 17-NOV-2000: 2000US-0249218.
 PR 17-NOV-2000: 2000US-0249244.
 PR 17-NOV-2000: 2000US-0249245.
 PR 17-NOV-2000: 2000US-0249264.
 PR 17-NOV-2000: 2000US-0249265.
 PR 17-NOV-2000: 2000US-0249297.
 PR 17-NOV-2000: 2000US-0249299.
 PR 17-NOV-2000: 2000US-0249300.
 PR 01-DEC-2000: 2000US-0250391.
 PR 01-DEC-2000: 2000US-0251160.
 PR 05-DEC-2000: 2000US-0251030.
 PR 05-DEC-2000: 2000US-0251988.
 PR 05-DEC-2000: 2000US-0256719.
 PR 06-DEC-2000: 2000US-0251479.
 PR 08-DEC-2000: 2000US-0251856.
 PR 08-DEC-2000: 2000US-0251868.
 PR 08-DEC-2000: 2000US-0251869.
 PR 08-DEC-2000: 2000US-0251988.
 PR 08-DEC-2000: 2000US-0251990.
 PR 11-DEC-2000: 2000US-0254097.
 PR 05-JAN-2001: 2001US-0259578.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 FI Rosen CA, Barash SC, Ruben SM;
 XX WPI: 2001-541565/60.
 DR P-PSDB: ABB15874.
 DR

XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 PS
 XX Claim 1: SEQ ID NO 1207; 1701bp + Sequence Listing; English.
 CC
 CC The invention relates to novel genes (HMA11004-ABA21534) and proteins
 CC (ABR14678-ABR18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at http://wipo.int/pub/publicsh-w_pcl_sequences.
 XX
 SO Sequence 236 BP; 69 A; 64 C; 39 G; 64 T; 0 other;
 Query Match 60.9%; Score 14; PH 22; Length 236;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 GCAATCTTACTTT 17
 Db 43 GCAATCTTACTTT 30
 RESULT 10
 ABR76032
 ID ABR76032.standard; DNA: 660 bp.
 XX
 AC ABR76032;
 XX
 DT 13-AUG-2002 (first entry)
 DE Bacillus licheniformis genomic sequence tag (GST) #3123.
 DE Bacillus licheniformis genomic sequenced tag; GST;
 KW differential gene expression; genomic sequenced tag; GST;
 KW altered culture condition; environmental stress;
 KW physiological provocation; ds.
 XX
 US Bacillus licheniformis.
 PN WO200229113-A2.
 PD 11-APR-2002.
 XX
 PF 05-OCT-2001; 2001WO-US31437.
 XX
 PR 06-OCT-2000; 2000US-0680598.
 PR 27-MAR-2001; 2001US-279526P.
 PA (NOVO) NOVOZYMES BIOTECH INC.
 PA (NOVO) NOVOZYMES AS.
 PI Berka R, Clausen IG;
 XX
 DR WPI: 2002-416684/44.
 XX
 PT Monitoring differential expression of several genes in first Bacillus
 PT cell relative to expression of same genes in one or more second
 PT Bacillus cells, by using substrate containing Bacillus genomic
 PT sequenced tag array -

XX PS Claim 4: SEQ ID NO 3323: 200pp: English.

XX CC The invention describes a method of monitoring differential expression of

CC genes in a first *Bacillus* cell relative to expression of the genes in

CC other *Bacillus* cells, comprising hybridizing labelled nucleic acid probes

CC isolated from *Bacillus* cells to a substrate containing array of *Bacillus*

CC genomic sequenced tags (GST), examining the array, and determining

CC relative gene expression by an observed hybridisation reporter signal of

CC a spot in the array. The method is useful for measuring the expression of

CC genes in a first *Bacillus* cell relative to expression of the same genes

CC in one or more second *Bacillus* cells. The method is useful for monitoring

CC global expression of several genes from a *Bacillus* cell, discovering new

CC genes, identifying possible functions of unknown open reading frames and

CC monitoring gene copy number variation and stability. Monitoring changes

CC in expression of genes may be used to provide a representation of the way

CC in which *Bacillus* cells adapt to changes in culture conditions,

CC environmental stress or other physiological provocation. Extensive

CC follow-up characterisation is unnecessary, when one spot on an array

CC equals one gene or one open reading frame, since sequence information is

CC available. This sequence represents a genomic sequence tag (GST) used in

CC the method of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WPI

CC at

CC ftp://ipo.int/pub/published_pct_sequences.

CC

SQ Sequence 660 BP: 224 A; 114 C; 137 G; 185 T; 0 other;

XX

Query Match 60.9%; Score 14; DB 24; Length 660;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 AATCTACTTCG 19

DB 371 AATCTACTTCG 384

RESULT 11

ID AB047496/5

XX AC AB047496 standard; DNA: 753 BP.

XX AC AB047496:

XX

DT 12-JUL-2002 (first entry)

XX

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 34087.

XX

KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;

KW SNP; cell differentiation; ds.

XX

OS Homo sapiens.

XX

PN W0200218632-A2.

XX

PD 07-MAR-2002.

XX

PF 01-SEP-2001; 2001WO-EP10074.

XX

PR 01-SEP-2000; 2000DE-1043826.

XX

PR 05-SEP-2000; 2000DE-1044543.

XX

PA (EPIC-) EPIGENOMICS AG.

XX

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX

DR WPI: 2002-371829/40.

XX

PT Determining the degree of cytosine methylation in genomic DNA, useful

PT for diagnosis and prognosis, comprises selective hybridization of

PT amplicons from chemically treated DNA

XX PS Claim 12: 56pp + Sequence Listing: 56pp; German.

XX CC This invention describes a novel method for determining the degree of

CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a

CC genomic sample of DNA. The sample is treated chemically to convert

CC cytosine (C) but not methylated C, to uracil, then part of the genomic

CC DNA that contains the target C is amplified to form a labeled amplicon.

CC The amplicon is hybridised to two classes, each with at least one

CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers

CC and the degree of hybridisation to both classes is determined from the

CC label on the amplicon. From the ratio of labels hybridised to the two

CC classes of oligomers, the degree of methylation is calculated. The method

CC is used: (i) for diagnosis and/or prognosis of side effects of

CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders

CC of the central nervous, cardiovascular, gastrointestinal and respiratory

CC systems etc., particularly by detecting mutations or single nucleotide

CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue

CC types and for investigating cell differentiation. The method allows the

CC methylation status of many C residues to be determined simultaneously.

CC AB013410-AB054121 represent genomic DNA sequences used to illustrate the

CC method for determining the degree of cytosine methylation described in

CC the disclosure of the invention.

CC

SQ Sequence 753 BP: 110 A; 77 C; 276 G; 290 T; 0 other;

XX

Query Match 60.9%; Score 14; DB 24; Length 753;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 TCTTACTTCGCTC 22

DB 56 TCTTACTTCGCTC 43

RESULT 12

ID AB047497

XX AC AB047497 standard; DNA: 753 BP.

XX AC AB047497:

XX

DT 12-JUL-2002 (first entry)

XX

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 34088.

XX

KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;

KW SNP; cell differentiation; ds.

XX

OS Homo sapiens.

XX

PN W0200218632-A2.

XX

PD 07-MAR-2002.

XX

PF 01-SEP-2001; 2001WO-EP10074.

XX

PR 01-SEP-2000; 2000DE-1043826.

XX

PR 05-SEP-2000; 2000DE-1044543.

XX

PA (EPIC-) EPIGENOMICS AG.

XX

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX

DR WPI: 2002-371829/40.

XX

PT Determining the degree of cytosine methylation in genomic DNA, useful

PT for diagnosis and prognosis, comprises selective hybridization of

PT amplicons from chemically treated DNA

XX

PS Claim 12: 56pp + Sequence Listing: 56pp; German.

CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

SO Sequence 753 BP: 290 A: 276 C: 77 G: 110 T: 0 other:

Query Match 60.9%; Score 14; DB 24; Length 753;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 TCTTACTTTCGCTC 22

DB 698 TCTTACTTTCGCTC 711

RESULT 13

ABQ23634/C

ID ABQ23634 standard; DNA: 767 BP.

AC ABQ23634:

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 10225.

XX Human; cytosine methylation: 5'-CpG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;

KW SNP; cell differentiation; ds.

OS Homo sapiens.

XX WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EPI0074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

PA (EPIC-) EPICENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

DR WPI: 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of
XX amplicons from chemically treated DNA

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one

CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

SO Sequence 767 BP: 283 A: 74 C: 318 G: 92 T: 0 other:

Query Match 60.9%; Score 14; DB 24; Length 767;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 TCTTACTTTCGCTC 22

DB 760 TCTTACTTTCGCTC 747

RESULT 14

ABQ23635

ID ABQ23635 standard; DNA: 767 BP.

AC ABQ23635:

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 10226.

XX Human; cytosine methylation: 5'-CpG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;

KW SNP; cell differentiation; ds.

OS Homo sapiens.

XX WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EPI0074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

PA (EPIC-) EPICENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

DR WPI: 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of
XX amplicons from chemically treated DNA

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one

CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC AB013410-AB054121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

XX
 SQ Sequence 767 BP; 92 A; 318 C; 74 G; 283 T; 0 other;

Query Match 60.9%; Score 14; DB 24; Length 767;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 TCTTACTTTCGCTC 22
 |||||
 DB 8 TCTTACTTTCGCTC 21

RESULT 15
 AAS86152/C
 ID AAS86152 standard; cDNA; 791 BP.
 XX
 AC AAS86152;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #21956.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YF;
 XX
 DR WPI: 2001-639362/73.
 DR P-PSDB: ABG21965.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 PS Claim 1; SEQ ID NO 21956; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPD
 CC at ftp.wipo.int/pub/published_pcl_sequences.

XX
 SQ Sequence 791 BP; 165 A; 202 C; 273 G; 151 T; 0 other;

Query Match 60.9%; Score 14; DB 23; Length 791;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 CTTTACTTTCGCTCC 23
 |||||
 DB 260 CTTTACTTTCGCTCC 247

Search completed: January 11, 2003, 23:01:21
 Job time : 314 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2003, 23:02:06 : Search time 2196 Seconds
(without alignments)
169.625 Million cell updates/sec

Title: SEQ1-T-AT-12347_COPY_12336_12358

Perfect score: 23
Sequence: 1 ggggcaaatcttacttcgctcc 23

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 12

Total number of hits satisfying chosen parameters: 8301

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

EST:
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estlov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrl:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	78.3	696	17	AC058557 Pan trogl
2	17	73.9	667	17	AZ305659 1M0006K24
3	16	69.6	391	12	BF426081 sr93f08.Y
4	16	69.6	591	17	AZ414906 1M0189F05
5	16	69.6	785	13	BJ141026 BJ141026
6	16	69.6	842	12	BF700342 BF700342

7	16	69.6	856	17	BH729616
8	16	69.6	1065	12	BE966142
9	15	65.2	101	9	AA210160
10	15	65.2	275	9	AA233037
11	15	65.2	276	9	AA715557
12	15	65.2	381	9	AI823498
13	15	65.2	384	9	AI991677
14	15	65.2	384	17	AC621595
15	15	65.2	404	14	T94015
16	15	65.2	405	9	AA714004
17	15	65.2	411	17	AZ946670
18	15	65.2	413	9	AA648991
19	15	65.2	416	9	AA516472
20	15	65.2	420	10	AA305385
21	15	65.2	422	9	AI990860
22	15	65.2	428	9	AI394056
23	15	65.2	437	10	AA131443
24	15	65.2	438	9	AI363505
25	15	65.2	443	9	AA139399
26	15	65.2	444	9	AA511395
27	15	65.2	448	9	AI651000
28	15	65.2	448	17	B54748
29	15	65.2	455	9	AA048668
30	15	65.2	463	9	AA831012
31	15	65.2	464	10	AA452869
32	15	65.2	477	17	AZ699818
33	15	65.2	480	9	A1765556
34	15	65.2	481	17	AA563015
35	15	65.2	486	10	BE082986
36	15	65.2	490	9	AA274671
37	15	65.2	499	12	BC586254
38	15	65.2	500	10	AA890856
39	15	65.2	501	17	BH379151
40	15	65.2	507	9	AA412300
41	15	65.2	513	17	AZ727592
42	15	65.2	518	9	AA874510
43	15	65.2	520	9	AI812790
44	15	65.2	524	14	BQ482637
45	15	65.2	545	17	AZ892283

ALIGNMENTS

RESULT 1	AG058557/c	696 bp	DNA	linear	GSS 02-NOV-2001
LOCUS	Pan troglodytes DNA, clone: pth-045H18.F, genomic survey sequence.				
DEFINITION	AG058557				
ACCESSION	AG058557.1	GI:16596018			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM	Pan troglodytes male lymphoblast DNA, clone: lib:PTH Chimpanzee Male				
REFERENCE	BAC library clone: pth-045H18.F.				
AUTHORS	Pan troglodytes				
TITLE	Totoki, Y., Watanabe, H. and Sakaki, Y.				
JOURNAL	BAC end sequences of library PTH				
AUTHORS	Unpublished				
TITLE	2 (bases 1 to 696)				
JOURNAL	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.				
COMMENT	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpanse@gsc.riken.go.jp, URL: http://npg.gsc.riken.go.jp/), Tel: 81-45-503-9111, Fax: 81-45-503-9170				
	Clones are derived from the chimpanzee BAC library PTH. This BAC end was generated during the Rad process and may have higher chance of clone tracking errors.				

PRIMERS

Sequencing: -21M13
LIBRARY

Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI.

FEATURES

Source

1..696
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-045H18.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 164 a 206 c 168 g 156 t 2 others
ORIGIN

Query Match 78.3% Score 18; DB 17; Length 696;
Best local Similarity 100.0%; Pred. No. 2.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGCGCAATCTTACTTCG 19
|||||
Db 387 GGCGCAATCTTACTTCG 370

RESULT 2
A2305659 667 bp DNA linear GSS 29-SEP-2000
LOCUS

DEFINITION
IM0006K24F Mouse 10kb plasmid UGCG1M library Mus musculus genomic
clone UGCG1M0006K24 F. DNA sequence.

ACCESSION
A2305659
KEYWORDS
GSS.

SOURCE
house mouse.

ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 667)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D. Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)

JOURNAL
COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
Bldg., USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 std error: 0.00
Plate: 0006 row: K column: 24
Seq primer: CGTTGTAAACGACGGCCACT
Class: plasmid ends
High quality sequence stop: 667.

FEATURES

Source

1..667
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCG1M0006K24"
/clone_lib="Mouse 10kb plasmid UGCG1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PMD42ny. Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a

BASE COUNT 164 a 155 c 151 g 197 t
ORIGIN

Query Match 73.9% Score 17; DB 17; Length 667;
Best local Similarity 100.0%; Pred. No. 9.8;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGCGCAATCTTACTTCG 18
|||||
Db 342 GGCGCAATCTTACTTCG 358

RESULT 3
BF426081 391 bp mRNA linear EST 06-DEC-2001
LOCUS

DEFINITION
sr93f08.y1 Gm-c1047 glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-c1047-1024 5', mRNA sequence.

ACCESSION
BF426081
KEYWORDS
EST.

SOURCE
soybean.

ORGANISM
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
REFERENCE
1 (bases 1 to 391)
Shoemaker,R., Keim,P., Vodka,L., Erpelting,J., Coryell,V., Khanna
A., Bolla,R., Marras,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wyle,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
Y., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk
R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
R., Waterston,R. and Wilson,R.
Public Soybean EST project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST project
Public Soybean EST project
Washington University School of Medicine
4444 Forest Park Parkway, box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: RegGen, Invitrogen Corp. 2136
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4763 or contact via email: couer@reggen.com
Insert length: 712 std error: 0.00
High quality sequence stop: 378.

FEATURES

Source

1..391
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1047-1024"
/clone_lib="Gm-c1047"
/tissue_type="immature leaves (unfurled trifoliolate) of
greenhouse grown plants"
/dev_stage="2 week old"
/lab_host="DH10B"
/note="Vector: pSPORT1, Site_1: NotI; Site_2: SalI; This
cDNA library was constructed from mRNA isolated from

immature leaves (unfurled trifoliate) of greenhouse grown plants that were 2 week old. The library was prepared using the Life Technologies pSPucript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly(dT) sequence with a NotI restriction site. SalI linker adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli Electromax DH10B host cells. This library was constructed in the laboratory of Dr. Lila Vodkin by Anu Khanna at the University of Illinois at Urbana-Champaign (email l-vodkin@uiuc.edu)

BASE COUNT 125 a 61 c 80 g 125 t
ORIGIN

Query Match 69.6%; Score 16; DB 12; Length 391;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCAATCTTACTT 16
|||||
DB 59 GGGCAATCTTACTT 74

RESULT 4

AZ414906/ 591 bp DNA linear GSS 03-OCT-2000
LOCUS 1M0189F05R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0189F05 R, DNA sequence.
ACCESSION AZ414906
VERSION AZ414906.1 GI:10538919
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 591)
AUTHORS Dunn, J.D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.
Moused whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std error: 0.00
Plate: 0189 row: F column: 05
Seq primer: CACACGAGAAACGCTATACAC
Class: plasmid ends
High quality sequence stop: 591.

FEATURES

Location/Qualifiers
1..591
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0189F05"
/clone_1lb="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42ny; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (911473211419b/AP129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 181 a 122 c 134 g 154 t
ORIGIN

Query Match 69.6%; Score 16; DB 17; Length 591;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGCATCTTACTTTC 18
|||||
DB 551 GGCATCTTACTTTC 536

RESULT 5

BJ141026/ 785 bp mRNA linear EST 23-JAN-2002
LOCUS BJ141026 unpublished oligo-capped cDNA library, C. elegans L1 stage
DEFINITION Caenorhabditis elegans cDNA clone yk1165b07.3, mRNA sequence.
ACCESSION BJ141026
VERSION BJ141026.1 GI:18301192
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans

REFERENCE Fukuoka, Y., Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae
1 (bases 1 to 785)
AUTHORS Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y., and Suano, S.
A complementary view of the C. elegans genome

JOURNAL Unpublished (2002)
COMMENT Contact: Tadasi Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-554-81-6856
Fax: 81-554-81-6855
Email: lshin@genes.nig.ac.jp.

Location/Qualifiers
1..785

FEATURES

Location/Qualifiers
1..785
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk1165b07"
/clone_1lb="unpublished oligo-capped cDNA library, C. elegans L1 stage"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/note="The AD-wrmcDNA library was generated with poly(A)+ RNA isolated from both hermaphrodite and male N2 worms of all larval stages, embryos, adults and dauers and the subsequent generation of cDNAs by poly(A) priming. The cDNAs were cloned into pPC86"
BASE COUNT 193 a 155 c 160 g 276 t 1 others
ORIGIN

Query Match 69.6%; Score 16; DB 13; Length 785;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CGCAATCTTACTTTC 18
 |||||
 Db 53 CGCAATCTTACTTTC 38

RESULT 6
 BF700342/c 842 bp mRNA linear EST 22-DEC-2000
 DEFINITION 60127771F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4284856 5',
 mRNA sequence.
 ACCESSION BF700342
 VERSION BF700342.1 GI:11985750
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 842)
 NIH-MGC http://mgi.ncl.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapds-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 plate: LICM118 row: m column: 17
 High quality sequence stop: 93.
 Location/Qualifiers
 1..842
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4284856"
 /clone_1id="NIH_MGC_56"
 /tissue_type="primitive neuroectoderm"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site:1:
 SfiI (ggcgccgcgcgc); Site:2: SfiI (ggcgccgcgcgc);
 Double-stranded cDNA was prepared from cell line RNA. 5'
 and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGGCCATTATGCGC-3' and 3' adaptor
 sequence: 5'-ATTCCTAGAGCGCGCGCGCGCATG-dT(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA)."

BASE COUNT 233 a 138 c 213 g 257 t 1 others

ORIGIN

Query Match 69.6%; Score 16; DB 12; Length 842;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 ATCTACTTTCGCTCC 23
 |||||
 Db 123 ATCTACTTTCGCTCC 108

RESULT 7
 BH729616 856 bp DNA linear GSS 20-FEB-2002
 BH729616
 LOCUS BOHMX08TR BO_2_3_KB Brassica oleracea genomic clone BOHMX08, DNA
 DEFINITION
 sequence.
 ACCESSION BH729616
 VERSION BH729616.1 GI:18835011
 KEYWORDS GSS.
 SOURCE Brassica oleracea.
 ORGANISM Brassica oleracea

REFERENCE
 1 (bases 1 to 856)
 Town, C.D., Van Aken, S., Uterback, T. and Fraser, C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 JOURNAL Unpublished (2001)
 COMMENT Other GSSs: BOHMX08TF
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: 7R
 Class: Sheared ends.
 Location/Qualifiers
 1..856
 /organism="Brassica oleracea"
 /strain="T01000DH3"
 /db_xref="taxon:3712"
 /clone="BOHMX08"
 /clone_1id="BO_2_3_KB"
 /note="Vector: pBIO1; Site:1: BstXI; 2-4 kb sheared
 genomic DNA inserted into pBIO1 using BstXI linkers"
 BASE COUNT 286 a 157 c 118 g 295 t

ORIGIN

Query Match 69.6%; Score 16; DB 17; Length 856;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AATCTACTTTCGCT 21
 |||||
 Db 46 AATCTACTTTCGCT 61

RESULT 8
 BE966142/c 1065 bp mRNA linear EST 14-DEC-2000
 LOCUS 601660070R1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3905825 3',
 mRNA sequence.
 ACCESSION BE966142
 VERSION BE966142.2 GI:11771240
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 1065)
 NIH-MGC http://mgi.ncl.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT On Oct 3, 2000 this sequence version replaced gi:10576847.
 Contact: Robert Strausberg, Ph.D.
 Email: cgapds-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Lite Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 plate: LICW700 row: l column: 18
 High quality sequence stop: 317.
 Location/Qualifiers
 1..1065
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3905825"
 /clone_1id="NIH_MGC_71"
 /tissue_type="telomysarcoma"

/lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 Kb."
 BASE COUNT 427 a 173 c 265 g 200 t
 ORIGIN

Query Match 69.6%; Score 16; DB 12; Length 1065;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 8 ATCTTACTTCGCTCC 23
 |||||||
 Db 868 ATCTTACTTCGCTCC 853

RESULT 9
 AA210160/c 101 bp mRNA linear EST 19-FEB-1997
 LOCUS mu40b04.r1 Soares_thymus_2NBMt Mus musculus cDNA clone IMAGE:641839
 DEFINITION
 5', mRNA sequence.
 ACCESSION AA210160
 VERSION AA210160.1 GI:1808535
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 101)
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)
 CONTACT: Maria M/Mouse EST Project
 WASHU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MG1:393831
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 87.

FEATURES
 source Location/Qualifiers
 1..101

/organism="Mus musculus"
 /strain="C57Bl/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:641839"
 /clone_lib="Soares_thymus_2NBMt"
 /sex="male"
 /tissue_type="thymus"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Vector: pT7T3D-Pac (pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer 15',
 TGTTCACATCGAGTGGAGCGCGCGTGTGTGTGTGTGTGTGTGTGT
 3'); double stranded cDNA was ligated to Eco RI adaptors
 (pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. RNA
 provided by Dr. Bertrand Jordan. Library went through two
 rounds of normalization, and was constructed by Bento
 Soares and M. Fatima Bonaldo."
 BASE COUNT 28 a 10 c 26 g 37 t
 ORIGIN

Query Match 65.2%; Score 15; DB 9; Length 101;

Best Local Similarity 100.0%; Pred. No. 90;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 4 GCAATCTTACTTTC 18
 |||||||
 Db 66 GCAATCTTACTTTC 52

RESULT 10
 AA233037/c 275 bp mRNA linear EST 06-AUG-1997
 LOCUS z146c09.r1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:666444
 DEFINITION
 5', mRNA sequence.
 ACCESSION AA233037
 VERSION AA233037.1 GI:1856030
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 275)
 Hillier,L., Allen,M., Bowles,M., Dubuque,T., Geisel,S., Jost,S.,
 Kucaba,T., Lacy,M., Le,M., Lennon,G., Marra,M., Martin,J., Moore,B.,
 Schellenberg,K., Steptoe,M., Tan,F., Theising,B., Wylie,T., Wylie,
 T., Waterston,R. and Wilson,R.
 WashU-Merck EST Project 1997
 Unpublished (1997)
 CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 2067 Std Error: 0.00
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 258.

FEATURES
 source Location/Qualifiers
 1..275

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /db_xref="IMAGE:666448"
 /clone="IMAGE:666448"
 /clone_lib="Soares_NHMPu_S1"
 /tissue_type="Pooled human melanocyte, fetal heart, and
 pregnant uterus"
 /lab_host="DH10B"
 /note="Organ: mixed (see below); Vector: pT7T3D-Pac
 (pharmacia) with a modified polylinker; Site_1: Not I;
 Site_2: Eco RI; Equal amounts of plasmid DNA from three
 normalized libraries (melanocyte 2bBM, pregnant uterus
 NBHPu, and fetal heart NBHPu) were mixed, and ss circles
 were made in vitro. Following HAP purification, this DNA
 was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from pools of
 5,000 clones made from the same 3 libraries. The pools
 consisted of I.M.A.G.E. clones 260232-265223,
 340488-345479, and 484488-489479."
 BASE COUNT 92 a 44 c 56 g 83 t
 ORIGIN

Query Match 65.2%; Score 15; DB 9; Length 275;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 GCAATCTTACTTT 17
 |||||||
 Db 107 GCAATCTTACTTT 93

RESULT 11
 AA15557/c

LOCUS AA715557 276 bp mRNA linear EST 22-JAN-1998
 DEFINITION nv53h10.r1 NCI_CGAP_Ew1 Homo sapiens cDNA clone IMAGE:1233571, mRNA
 sequence.
 ACCESSION AA715557
 VERSION AA715557
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 276)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabps-r@mail.nih.gov
 Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
 Ph.D.
 CDNA Library Preparation: David B. Krizman, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/dbp/image/image.html
 Insert Length: 404 Std Error: 0.00
 Seq primer: -28m13 rev1 ET from Amersham.
 Location/Qualifiers
 source
 1. 276
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1233571"
 /clone_1ib="NCI_CGAP_Ew1"
 /tissue_type="Ewing's sarcoma"
 /lab_host="DH10B"
 /note="Vector: PAMP10; mRNA made from Ewing's sarcoma,
 cDNA made by oligo-dT priming. Non-directionally cloned.
 Size-selected on agarose gel, average insert size 600 bp.
 Reference: Krizman et al. (1996) Cancer Research
 56:5380-5383."

BASE COUNT 63 a 61 c 70 g 82 t
 ORIGIN
 Query Match 65.2%; Score 15; DB 9; Length 276;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGGCAATCTTACTT 16
 Db 232 GGGCAATCTTACTT 218

RESULT 12
 LOCUS A1823498 381 bp mRNA linear EST 20-DEC-1999
 DEFINITION wh54909.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2384608 3',
 mRNA sequence.
 ACCESSION A1823498
 VERSION A1823498
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 381)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabps-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.

LOCUS AA715557 276 bp mRNA linear EST 22-JAN-1998
 DEFINITION nv53h10.r1 NCI_CGAP_Ew1 Homo sapiens cDNA clone IMAGE:1233571, mRNA
 sequence.
 ACCESSION AA715557
 VERSION AA715557
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 276)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabps-r@mail.nih.gov
 Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
 Ph.D.
 CDNA Library Preparation: David B. Krizman, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/dbp/image/image.html
 Insert Length: 404 Std Error: 0.00
 Seq primer: -28m13 rev1 ET from Amersham.
 Location/Qualifiers
 source
 1. 276
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1233571"
 /clone_1ib="NCI_CGAP_Ew1"
 /tissue_type="Ewing's sarcoma"
 /lab_host="DH10B"
 /note="Vector: PAMP10; mRNA made from Ewing's sarcoma,
 cDNA made by oligo-dT priming. Non-directionally cloned.
 Size-selected on agarose gel, average insert size 600 bp.
 Reference: Krizman et al. (1996) Cancer Research
 56:5380-5383."

BASE COUNT 63 a 61 c 70 g 82 t
 ORIGIN
 Query Match 65.2%; Score 15; DB 9; Length 276;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGGCAATCTTACTT 16
 Db 232 GGGCAATCTTACTT 218

RESULT 12
 LOCUS A1823498 381 bp mRNA linear EST 20-DEC-1999
 DEFINITION wh54909.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2384608 3',
 mRNA sequence.
 ACCESSION A1823498
 VERSION A1823498
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 381)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabps-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.

/tissue_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)"
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 92 a 97 c 77 g 117 t 1 others

ORIGIN

Query Match 65.2%; Score 15; DB 9; Length 384;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGCAATCTTACTT 16
 Db 304 GGCAATCTTACTT 290

RESULT 14
 A0621595/c
 LOCUS HS_3041.A1.A05.MR CIT Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone plate-3041 Col=9 Row=A, DNA sequence.
 ACCESSION A0621595
 VERSION A0621595.1 GI:5083987
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 384)
 Mahatras,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Kellar,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 High quality sequence stop: 384.
 Location/Qualifiers
 1. 384
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_plate="3041 Col=9 Row=A"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: Phe10BAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 128 a 50 c 90 g 116 t

ORIGIN

Query Match 65.2%; Score 15; DB 17; Length 384;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AATCTTACTTTCGC 20
 Db 54 AATCTTACTTTCGC 40

RESULT 15
 T94015
 LOCUS T94015 404 bp mRNA
 DEFINITION yez9g08.r1 Stratogene lung (#937210) Homo sapiens cDNA clone
 SEQUENCE:119198 5' similar to contains TARI repetitive element; , mRNA
 1 (bases 1 to 404)

ACCESSION T94015
 VERSION T94015.1 GI:727503
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 404)
 Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiappelli,B., Chissole,S., Dietrich,N., Dubouque,T., Favello,A., Gish,M., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,D., Trevisan,E., Underwood,K., Wohlmann,P., Waterson,R., Wilson,R. and Merris,M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)

TITLE
 JOURNAL
 MEDLINE
 COMMENT
 97044478
 Contact: Wilson RK
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 539
 High quality sequence stop: 271.
 This clone is available royalty-free through LNL, the IMAGE Consortium (imgcimage.lnl.gov) for further information.
 Insert Length: 539 Std Error: 0.00
 Seq primer: M13RP1
 High quality sequence stop: 271.
 Location/Qualifiers
 1. 404
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /db_xref="IMAGE:119198"
 /clone_plate="119198"
 /clone_lib="Stratogene lung (#937210)"
 /sex="male"
 /dev_stage="72 years"
 /lab_host="Solk cells (kanamycin resistant)"
 /note="Organ: lung; Vector: philescrip SK-; Site: J; EROI
 ; Site: 2; XHO1; Cloned unidirectionally. Primer: oligo
 dt, normal lung. Average insert size: 1.0 kb; Uni-ZAP AR
 Vector: -5' adaptor sequence: 5' GATTTCGACGACG 3' -3'
 adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTT 3"

BASE COUNT 109 a 88 c 86 g 116 t 5 others

ORIGIN

Query Match 65.2%; Score 15; DB 14; Length 404;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGCAATCTTACTT 17
 Db 319 GGCAATCTTACTT 333

Search completed: January 12, 2003, 00:27:25
 Job time : 2208 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2003, 23:01:32 ; Search time 2901 Seconds

(without alignments)
230.736 Million cell updates/sec

Title: SF01-T-AT-12347_COPY_12336_12358

Perfect score: 23

Sequence: 1 9999caactcttactcttctcc 23

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 12

Total number of hits satisfying chosen parameters: 15901

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenFtbl:*

1: gb_ba:*

2: gb_hcg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_hcg_hum:*

31: em_hcg_inv:*

32: em_hcg_other:*

33: em_hcg_mus:*

34: em_hcg_pln:*

35: em_hcg_rod:*

36: em_hcg_mam:*

37: em_hcg_vrt:*

38: em_sy:*

39: em_hgtgo_hum:*

40: em_hgtgo_mus:*

41: em_hgtgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	DB ID	Description
1	23	100.0	42479	AC074203
2	23	100.0	115027	AP001480
3	23	100.0	175582	AP001481
4	17	73.9	63669	AC109505
5	17	73.9	63669	AC109505
6	17	73.9	177182	AC124530
7	17	73.9	202953	AC130947
8	17	73.9	218369	AC122351
9	17	73.9	235301	AC124772
10	17	73.9	239619	AC124601
11	16	69.6	700	AX183106
12	16	69.6	2238	AF086808
13	16	69.6	2558	BC030525
14	16	69.6	24102	AC004649
15	16	69.6	33734	CEC05C9
16	16	69.6	34721	HS315G5
17	16	69.6	140287	HSJ407E4
18	16	69.6	146797	AC034220
19	16	69.6	148948	AF005518
20	16	69.6	152996	AC013719
21	16	69.6	157747	AC067915
22	16	69.6	163660	AC046165
23	16	69.6	164343	AC024593
24	16	69.6	185867	AC074311
25	16	69.6	187486	AC079840
26	16	69.6	191397	AC023861
27	16	69.6	191397	AC023861
28	16	69.6	270150	AF006639
29	15	65.2	276	AX163708
30	15	65.2	381	AX163719
31	15	65.2	384	AX163721
32	15	65.2	405	AX163707
33	15	65.2	413	AX163706
34	15	65.2	416	AX163705
35	15	65.2	422	AX163720
36	15	65.2	428	AX163713
37	15	65.2	437	AX163723
38	15	65.2	438	AX163712
39	15	65.2	448	AX163716
40	15	65.2	462	AX260411
41	15	65.2	464	AX164710
42	15	65.2	464	AX164724
43	15	65.2	480	AX164718
44	15	65.2	630	AX163699
45	15	65.2	928	AF004228

ALIGNMENTS

RESULT 1

AC074203

LOCUS AC074203 42479 bp DNA linear PK1 13-OCT-2000

DEFINITION Homo sapiens chromosome 22q11 clone cos4, complete sequence.

AC074203

AC074203.3 GI:9625348

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 42479)

Yang, L., Hu, P., Wu, H., Tapia-Paez, I. and Roe, B.A.

AUTHORS

Homo sapiens Chromosome 22q11 Cosmid cos4

JOURNAL

Unpublished

REMARK	This clone is a chimeric cosmid from a patient with t(11;22) translocation and it is from the derivative chromosome 11. There are sequences from both chromosome 11 and 22 in this cosmid				
REFERENCE	2 (bases 1 to 42479)				
AUTHORS	Yang, L., Hu, P., Wu, H., Tapia-Paez, I. and Roe, B.A.				
TITLE	Direct Submission				
JOURNAL	Submitted (18-JUL-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA				
REFERENCE	3 (bases 1 to 42479)				
AUTHORS	Yang, L., Hu, P., Wu, H., Tapia-Paez, I. and Roe, B.A.				
TITLE	Direct Submission				
JOURNAL	Submitted (03-AUG-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA				
REFERENCE	4 (bases 1 to 42479)				
AUTHORS	Yang, L., Hu, P., Wu, H., Tapia-Paez, I. and Roe, B.A.				
TITLE	Direct Submission				
JOURNAL	Submitted (13-OCT-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA				
COMMENT	On Aug 1, 2000 this sequence version replaced gi:7286164.				
FEATURES	Location/Qualifiers				
SOURCE	1. 42479				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/chromosome="22q11"				
	/clone="cos4"				
BASE COUNT	11501 a	9314 c	9563 g	12101 t	
ORIGIN					
Query Match	100.0%	Score 23:	DB 9:	Length 42479:	
Best Local Similarity	100.0%	Pred. No. 0.0089;			
Matches 23: Conservative	0:	Mismatches 0:	Indels 0:	Gaps 0:	
QY	1	GGGGCAATCTTACTTTGCTGC	23		
Db	31444	GGGGCAATCTTACTTTGCTGC	31466		
RESULT 2					
AP001480/c	Homo sapiens chromosome 11 clone CTC-227C10 map 11q, WORKING DRAFT				
LOCUS	AP001480	115027 bp	DNA	linear	HTG 30-MAY-2000
DEFINITION	Homo sapiens chromosome 11 clone CTC-227C10 map 11q, WORKING DRAFT				
ACCESSION	AP001480				
VERSION	AP001480.2 GI:8117338				
KEYWORDS	HTG; HTGS; PHASE1; HTGS; DRAFT.				
SOURCE	Homo sapiens DNA, clone:CTC-227C10.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
REFERENCE	1 (bases 1 to 115027)				
AUTHORS	Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.				
TITLE	Homo sapiens 115,027 genomic DNA of 11q				
JOURNAL	Published Only In Database (2000)				
REFERENCE	2 (bases 1 to 115027)				
AUTHORS	Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-MAR-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan (E-mail:hattori@gsc.riken.go.jp,				
COMMENT	URL:http://hgp.gsc.riken.go.jp/, tel:81-42-778-9923, Fax:81-42-778-9924)				
	On May 30, 2000 this sequence version replaced gi:7286164.				
	----- Genomic Center				
	Center: RIKEN Genomic Sciences Center (GSC)				
	Center code: RIKEN				
	Web site: http://hgp.gsc.riken.go.jp/				
	Contact: hattori@gsc.riken.go.jp				

```

----- Project Information -----
Center Project name: HumDrat11
Center Clone name: CTC-227C10

----- Summary Statistics -----
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amers39
Assembly program: Phrap; version 0.990329
Consensus quality: 107728 bases at least Q40
Consensus quality: 110501 bases at least Q30
Consensus quality: 11810 bases at least Q20
Insert size: 113327; sum-of-ctrlsigs
Quality coverage: 6.70x in Q20 bases; sum-of-ctrlsigs

NOTE: This is a 'working draft' sequence. It currently consists of
18 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

1      15681 contig of 15681 bp in length
15782      30275 contig of 14494 bp in length
30376      40528 contig of 10153 bp in length
40629      50205 contig of 9577 bp in length
50306      59617 contig of 9312 bp in length
59718      66750 contig of 7033 bp in length
66851      73847 contig of 6997 bp in length
73948      79644 contig of 5697 bp in length
79745      85653 contig of 5909 bp in length
85654      85753 gap of 100 bp
85754      91241 contig of 5488 bp in length
91242      91341 gap of 100 bp
91342      96419 contig of 5078 bp in length
96420      96519 gap of 100 bp
96520      100644 contig of 4125 bp in length
100645      100744 gap of 100 bp
100745      102923 contig of 2179 bp in length
102924      103023 gap of 100 bp

Sequence updated (26-May-2000).
NOTE: This is a 'working draft' sequence. It currently
consists of 18 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1      15681 contig of 15681 bp in length
15682      15781 gap of 100 bp
15782      30275 contig of 14494 bp in length
30276      40275 gap of 100 bp
40276      40528 contig of 10153 bp in length
40529      40628 gap of 100 bp
40629      50205 contig of 9577 bp in length
50206      50305 gap of 100 bp
50306      59617 contig of 9312 bp in length
59618      59717 gap of 100 bp
59718      66750 contig of 7033 bp in length
66751      66850 gap of 100 bp
66851      73847 contig of 6997 bp in length
73848      73947 gap of 100 bp
73948      79644 contig of 5697 bp in length
79645      79744 gap of 100 bp
79745      85653 contig of 5909 bp in length
85654      85753 gap of 100 bp
85754      91241 contig of 5488 bp in length
91242      91341 gap of 100 bp
91342      96419 contig of 5078 bp in length
96420      96519 gap of 100 bp
96520      100644 contig of 4125 bp in length
100645      100744 gap of 100 bp
100745      102923 contig of 2179 bp in length
102924      103023 gap of 100 bp

```

```

* 103024 105950: contig of 2927 bp in length
* 105951 106050: gap of 100 bp
* 106051 108938: contig of 2888 bp in length
* 108939 109038: gap of 100 bp
* 109039 111577: contig of 2539 bp in length
* 111578 111677: gap of 100 bp
* 111678 113487: contig of 1810 bp in length
* 113488 113587: gap of 100 bp
* 113588 115027: contig of 1440 bp in length.

```

FEATURES

```

source
1..115027
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="CTC-270C21"
1..15681
/misc_feature
/note="assembly-fragment"
15782..30275
/misc_feature
/note="assembly-fragment"
30376..40528
/misc_feature
/note="assembly-fragment"
40629..50205
/misc_feature
/note="assembly-fragment"
50306..59617
/misc_feature
/note="assembly-fragment"
59718..66750
/misc_feature
/note="assembly-fragment"
66851..73847
/misc_feature
/note="assembly-fragment"
73948..79644
/misc_feature
/note="assembly-fragment"
79745..85653
/misc_feature
/note="assembly-fragment"
85754..91241
/misc_feature
/note="assembly-fragment"
91342..96419
/misc_feature
/note="assembly-fragment"
96520..100644
/misc_feature
/note="assembly-fragment"
100745..102923
/misc_feature
/note="assembly-fragment"
103024..105950
/misc_feature
/note="assembly-fragment"
106051..108938
/misc_feature
/note="assembly-fragment"
109039..111577
/misc_feature
/note="assembly-fragment"
111678..113487
/misc_feature
/note="assembly-fragment"
113588..115027
misc_feature
/note="assembly-fragment"
BASE COUNT 30684 a 26526 c 26291 g 29826 t 1700 others
ORIGIN
Query Match 100.0%: Score 23; DB 2; Length 115027;
Best local Similarity 100.0%: Pred. No. 0.0076;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GGGCAATCTTACTTTCGCTC 23
|||||
Db 55098 GGGCAATCTTACTTTCGCTC 55076

```

```

RESULT 3
AP001481/c 175582 bp DNA linear HTG 30-MAY-2000
LOCUS Homo sapiens chromosome 11 clone CTC-270C21 map 11q, WORKING DRAFT
DEFINITION SEQUENCE, 29 unordered pieces.
ACCESSION AP001481
VERSION AP001481.2 GI:8117339
KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE Homo sapiens DNA, clone: CTC-270C21.

```

```

ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 175582)
Hattori, M., Ishii, K., Toyoda, A., Taylor, T. P., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
Homo sapiens 175,582 genomic DNA of 11q
2 (bases 1 to 175582)
Published Only in Database (2000)
Hattori, M., Ishii, K., Toyoda, A., Taylor, T. P., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (17-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail: hattori@gsc.riken.go.jp,
URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923,
Fax: 81-42-778-9924)
On May 30, 2000 this sequence version replaced gi:7288175.

```

COMMENT

```

----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: Humrall11
Center clone name: CTC-270C21
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 167630 bases at least Q40
Consensus quality: 167630 bases at least Q30
Insert size: 172782; sum-of-ontigs
Quality coverage: 4.96x in Q20 bases; sum-of-ontigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of
29 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs 'N', but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved.
1 16627 contig of 16627 bp in length
16728 30620 contig of 13893 bp in length
30721 42439 contig of 11719 bp in length
42540 51421 contig of 10882 bp in length
53522 64748 contig of 11227 bp in length
64849 76919 contig of 12071 bp in length
77020 85523 contig of 8504 bp in length
85624 94708 contig of 9085 bp in length
94809 104076 contig of 9286 bp in length
104177 112629 contig of 8453 bp in length
112730 119309 contig of 6580 bp in length
119410 126696 contig of 7287 bp in length
126797 132422 contig of 5626 bp in length
132523 137920 contig of 5398 bp in length
138021 142190 contig of 4170 bp in length
142291 145283 contig of 2993 bp in length
145384 148550 contig of 3167 bp in length
148651 151425 contig of 2775 bp in length
151526 155432 contig of 3907 bp in length
155533 158018 contig of 2486 bp in length
158119 160661 contig of 2543 bp in length
160762 163242 contig of 2481 bp in length
163343 165254 contig of 1912 bp in length
165355 167624 contig of 2270 bp in length
167725 169398 contig of 1674 bp in length
169499 171146 contig of 1648 bp in length
171247 173044 contig of 1738 bp in length
173145 174332 contig of 1188 bp in length
174433 175582 contig of 1150 bp in length

```

Sequence updated (26-May-2000)

* NOTE: This is a 'working draft' sequence. It currently consists of 29 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
1 16627: contig of 16627 bp in length
16628 16727: gap of 100 bp
16728 30620: contig of 13893 bp in length
30621 30720: gap of 100 bp
30721 42439: contig of 11719 bp in length
42440 42539: gap of 100 bp
42540 53421: contig of 10882 bp in length
53422 53521: gap of 100 bp
53522 64748: contig of 11227 bp in length
64749 64848: gap of 100 bp
64849 76919: contig of 12071 bp in length
76920 77019: gap of 100 bp
77020 85523: contig of 8504 bp in length
85524 94708: contig of 9085 bp in length
94709 94808: gap of 100 bp
94809 104076: contig of 9268 bp in length
104077 104176: gap of 100 bp
104177 112629: contig of 8453 bp in length
112630 112729: gap of 100 bp
112730 119309: contig of 6580 bp in length
119310 119409: gap of 100 bp
119410 126696: contig of 7287 bp in length
126697 126796: gap of 100 bp
126797 132422: contig of 5626 bp in length
132423 132522: gap of 100 bp
132523 137920: contig of 5398 bp in length
137921 138020: gap of 100 bp
138021 142190: contig of 4170 bp in length
142191 142290: gap of 100 bp
142291 145283: contig of 2993 bp in length
145284 145383: gap of 100 bp
145384 148550: contig of 3167 bp in length
148551 148650: gap of 100 bp
148651 151425: contig of 2775 bp in length
151426 151525: gap of 100 bp
151526 155432: contig of 3907 bp in length
155433 155532: gap of 100 bp
155533 158018: contig of 2486 bp in length
158019 158118: gap of 100 bp
158119 160661: contig of 2544 bp in length
160662 160761: gap of 100 bp
160762 163242: contig of 2481 bp in length
163243 163342: gap of 100 bp
163343 165254: contig of 1912 bp in length
165255 165354: gap of 100 bp
165355 167624: contig of 2270 bp in length
167625 167724: gap of 100 bp
167725 169398: contig of 1674 bp in length
169399 169498: gap of 100 bp
169499 171146: contig of 1648 bp in length
171147 171246: gap of 100 bp
171247 173044: contig of 1798 bp in length
173045 173144: gap of 100 bp
173145 174332: contig of 1188 bp in length
174333 174432: gap of 100 bp
174433 175582: contig of 1150 bp in length.

Location/Qualifiers
1. 175582
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="CFC-270C21"
```

```
misc_feature 1..16627
16728..30620
/misc_feature /note="assembly-fragment"
30721..42439
/misc_feature /note="assembly-fragment"
42540..53421
/misc_feature /note="assembly-fragment"
53522..64748
/misc_feature /note="assembly-fragment"
64849..76919
/misc_feature /note="assembly-fragment"
77020..85523
/misc_feature /note="assembly-fragment"
85624..94708
/misc_feature /note="assembly-fragment"
94809..104076
/misc_feature /note="assembly-fragment"
104177..112629
/misc_feature /note="assembly-fragment"
112730..119409
/misc_feature /note="assembly-fragment"
119410..126696
/misc_feature /note="assembly-fragment"
126797..132422
/misc_feature /note="assembly-fragment"
132523..137920
/misc_feature /note="assembly-fragment"
138021..142190
/misc_feature /note="assembly-fragment"
142291..145283
/misc_feature /note="assembly-fragment"
145384..148550
/misc_feature /note="assembly-fragment"
148651..151425
/misc_feature /note="assembly-fragment"
151526..155432
/misc_feature /note="assembly-fragment"
155533..158018
/misc_feature /note="assembly-fragment"
158119..160661
/misc_feature /note="assembly-fragment"
160762..163242
/misc_feature /note="assembly-fragment"
163343..165254
/misc_feature /note="assembly-fragment"
165355..167624

Query Match 100.0%; Score 24; H: 2; Length 175582;
Host Local Similarity 100.0%; Pred. No. 0.0071;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

UY 1 GGGGCAATCTTACCTTCGCTCC 23
|||||
Db 108106 GGGGCAATCTTACCTTCGCTCC 108084

RESULT 4
AC109505 63669 bp DNA linear HTG 04-FEB-2002
LOCUS Mus musculus clone RP23-155M20, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC109505
ACCESSION AC109505.1 GI:18483516
VERSION
KEYWORDS HTG; HTGS_PHASED.
SOURCE
ORGANISM Mus musculus
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
REFERENCE 1 (bases 1 to 63669)
AUTHORS Birren,B., Lincoln,L., Nusbaum,C. and Lander,E.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 63669)
```

AUTHORS

Birtten, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barne, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B., Chopel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kellis, C., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McKean, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenga, Y., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhian, P., Pierre, N., Pollara, V., Raymond, C., Rella, R., Rieback, M., Riley, R., Riso, C., Rogov, P., Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Strange-Thumann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (04-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Smit, A.F.A. & Green, P. (1996-1997)

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Web site: http://www.seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L20871

Center clone name: 155_M_20

* NOTE: This record contains 80 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
701 800: contig of 700 bp in length
801 1500: contig of 700 bp in length
1501 1600: gap of 100 bp
1601 2254: contig of 654 bp in length
2255 2354: gap of 100 bp
2355 3058: contig of 704 bp in length
3059 3158: gap of 100 bp
3159 3869: contig of 711 bp in length
3870 3969: gap of 100 bp
3970 4683: contig of 714 bp in length
4684 4783: gap of 100 bp
4784 5482: contig of 699 bp in length
5483 5582: gap of 100 bp
5583 6284: contig of 702 bp in length
6285 6384: gap of 100 bp
6385 7067: contig of 683 bp in length
7068 7167: gap of 100 bp
7168 7867: contig of 700 bp in length
7868 7967: gap of 100 bp
7968 8651: contig of 684 bp in length
8652 8751: gap of 100 bp
8752 9470: contig of 719 bp in length
9471 9570: gap of 100 bp
9571 10264: contig of 694 bp in length

10265 10364: gap of 100 bp
10365 11045: contig of 681 bp in length
11046 11145: gap of 100 bp
11146 11864: contig of 719 bp in length
11865 11964: gap of 100 bp
11965 12666: contig of 702 bp in length
12667 12766: gap of 100 bp
12767 13471: contig of 705 bp in length
13472 13571: gap of 100 bp
13572 14250: contig of 679 bp in length
14251 14350: gap of 100 bp
14351 15046: contig of 696 bp in length
15047 15146: gap of 100 bp
15147 15828: contig of 682 bp in length
15829 15928: gap of 100 bp
15929 16629: contig of 701 bp in length
16630 16729: gap of 100 bp
16730 17424: contig of 695 bp in length
17425 17524: gap of 100 bp
17525 18216: contig of 692 bp in length
18217 18316: gap of 100 bp
18317 19010: contig of 694 bp in length
19011 19110: gap of 100 bp
19111 19821: contig of 711 bp in length
19822 19921: gap of 100 bp
19922 20627: contig of 706 bp in length
20628 20727: gap of 100 bp
20728 21433: contig of 706 bp in length
21434 21533: gap of 100 bp
21534 22238: contig of 705 bp in length
22239 22338: gap of 100 bp
22339 23038: contig of 700 bp in length
23039 23138: gap of 100 bp
23139 23837: contig of 699 bp in length
23838 23937: gap of 100 bp
23938 24632: contig of 695 bp in length
24633 24732: gap of 100 bp
24733 25426: contig of 694 bp in length
25427 25526: gap of 100 bp
25527 26223: contig of 697 bp in length
26224 26323: gap of 100 bp
26324 27022: contig of 699 bp in length
27023 27122: gap of 100 bp
27123 27825: contig of 703 bp in length
27826 27925: gap of 100 bp
27926 28632: contig of 707 bp in length
28633 28732: gap of 100 bp
28733 29428: contig of 696 bp in length
29429 29528: gap of 100 bp
29529 30222: contig of 694 bp in length
30223 30322: gap of 100 bp
30323 31013: contig of 691 bp in length
31014 31113: gap of 100 bp
31114 31806: contig of 693 bp in length
31807 31906: gap of 100 bp
31907 32605: contig of 699 bp in length
32606 32705: gap of 100 bp
32706 33394: contig of 689 bp in length
33395 33494: gap of 100 bp
33495 34187: contig of 693 bp in length
34188 34287: gap of 100 bp
34288 34989: contig of 702 bp in length
34990 35089: gap of 100 bp
35090 35798: contig of 709 bp in length
35799 35898: gap of 100 bp
35899 36597: contig of 699 bp in length
36598 36697: gap of 100 bp
36698 37398: contig of 701 bp in length
37399 37498: gap of 100 bp
37499 38207: contig of 709 bp in length
38208 38307: gap of 100 bp
38308 39006: contig of 699 bp in length
39007 39106: gap of 100 bp

```

* 39107 39803: contig of 697 bp in length
* 39804 39903: gap of 100 bp
* 39904 40623: contig of 720 bp in length
* 40624 40723: gap of 100 bp
* 40724 41415: contig of 692 bp in length
* 41416 41515: gap of 100 bp
* 41516 42212: contig of 697 bp in length
* 42213 42312: gap of 100 bp
* 42313 43005: contig of 693 bp in length
* 43006 43105: gap of 100 bp
* 43106 43792: contig of 687 bp in length
* 43793 43892: gap of 100 bp
* 43893 44601: contig of 709 bp in length
* 44602 44701: gap of 100 bp
* 44702 45398: contig of 697 bp in length
* 45399 45498: gap of 100 bp
* 45499 46189: contig of 691 bp in length
* 46190 46289: gap of 100 bp
* 46290 46980: contig of 691 bp in length
* 46981 47080: gap of 100 bp
* 47081 47780: contig of 700 bp in length
* 47781 47880: gap of 100 bp
* 47881 48582: contig of 702 bp in length
* 48583 48682: gap of 100 bp
* 48683 49369: contig of 687 bp in length
* 49370 49469: gap of 100 bp
* 49470 50159: contig of 690 bp in length
* 50160 50259: gap of 100 bp
* 50260 50958: contig of 699 bp in length
* 50959 51058: gap of 100 bp
* 51059 51765: contig of 707 bp in length
* 51766 51865: gap of 100 bp
* 51866 52568: contig of 703 bp in length
* 52569 52668: gap of 100 bp
* 52669 53356: contig of 688 bp in length
* 53357 53456: gap of 100 bp
* 53457 54153: contig of 697 bp in length
* 54154 54253: gap of 100 bp
* 54254 54947: contig of 694 bp in length
* 54948 55047: gap of 100 bp

```

Query Match 73.9% Score 17; DB 2: Length 63669;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGGCAAACTTACTTTC 18
 Db 47984 GGGCAAACTTACTTTC 48000

RESULT 5
 AC109505/C 63669 bp DNA linear HTG 04-FEB-2002
 LOCUS Mus musculus clone RP23-155M20, LOW-PASS SEQUENCE SAMPLING.
 DEFINITION AC109505
 AC109505.1 GI:18483516
 VERSION HTG: HTGS_PHASE0.
 KEYWORDS Mus musculus.
 SOURCE Mus musculus.
 ORGANISM Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 63669)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Mus musculus, clone RP23-155M20
 JOURNAL Unpublished
 2 (bases 1 to 63669)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
 Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B.,
 Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., DeRellano, K., Dewar, K., Diaz, J., Dodge, S., Ferro, S.,
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

TITLE JOURNAL COMMENT

Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamel, A., Karatas, A., Kells, C., Labocque, K., Lamazares, R.,
 Landers, T., Lehoczek, J., Levine, R., Liu, G., Maclean, C.,
 MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
 McEwan, P., McKernan, K., Meldrum, J., Menes, L., Mihova, T.,
 Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Notbu, C.,
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
 Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
 Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
 Roselli, R., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S.,
 Severly, P., Spencer, R., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trifillo, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zaimoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (04-FEB-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBC
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu
 Project Information
 Center project name: L20871
 Center clone name: 155_M-20

NOTE: This record contains 40 individual
 sequencing reads that have not been assembled into
 contigs. Runs of N are used to separate the reads
 and the order in which they appear is completely
 arbitrary. Low-pass sequence sampling is useful for
 identifying clones that may be gene-rich and allows
 overlap relationships among clones to be deduced.
 However, it should not be assumed that this clone
 will be sequenced to completion. In the event that
 the record is updated, the accession number will
 be preserved.

```

1 700: contig of 700 bp in length
* 701 800: gap of 100 bp
* 801 11500: contig of 700 bp in length
* 1501 1600: gap of 100 bp
* 1601 2254: contig of 654 bp in length
* 2255 2354: gap of 100 bp
* 2355 3058: contig of 704 bp in length
* 3059 3158: gap of 100 bp
* 3159 3869: contig of 711 bp in length
* 3870 3969: gap of 100 bp
* 3970 4683: contig of 714 bp in length
* 4684 4783: gap of 100 bp
* 4784 5482: contig of 699 bp in length
* 5483 5582: gap of 100 bp
* 5583 6284: contig of 702 bp in length
* 6285 6384: gap of 100 bp
* 6385 7067: contig of 683 bp in length
* 7068 7167: gap of 100 bp
* 7168 7867: contig of 700 bp in length
* 7868 7967: gap of 100 bp
* 7969 8651: contig of 684 bp in length
* 8652 8751: gap of 100 bp
* 8752 9470: contig of 719 bp in length
* 9471 9570: gap of 100 bp
* 9571 10264: contig of 694 bp in length
* 10265 10364: gap of 100 bp
* 10365 11045: contig of 681 bp in length
* 11046 11145: gap of 100 bp
* 11146 11864: contig of 719 bp in length
* 11865 11964: gap of 100 bp
* 11965 12666: contig of 702 bp in length
* 12667 12766: gap of 100 bp

```



```

12767 13471: contig of 705 bp in length
13472 13571: gap of 100 bp
13572 14250: contig of 679 bp in length
14251 14350: gap of 100 bp
14351 15046: contig of 696 bp in length
15047 15146: gap of 100 bp
15147 15828: contig of 682 bp in length
15829 15928: gap of 100 bp
15929 16629: contig of 701 bp in length
16630 16729: gap of 100 bp
16730 17424: contig of 695 bp in length
17425 17524: gap of 100 bp
17525 18216: contig of 692 bp in length
18217 18316: gap of 100 bp
18317 19010: contig of 694 bp in length
19011 19110: gap of 100 bp
19111 19821: contig of 711 bp in length
19822 19921: gap of 100 bp
19922 20627: contig of 706 bp in length
20628 20727: gap of 100 bp
20728 21433: contig of 706 bp in length
21434 21533: gap of 100 bp
21534 22238: contig of 705 bp in length
22239 22338: gap of 100 bp
22339 23038: contig of 700 bp in length
23039 23138: gap of 100 bp
23139 23837: contig of 699 bp in length
23838 23937: gap of 100 bp
23938 24632: contig of 695 bp in length
24633 24732: gap of 100 bp
24733 25426: contig of 694 bp in length
25427 25526: gap of 100 bp
25527 26223: contig of 697 bp in length
26224 26323: gap of 100 bp
26324 27022: contig of 699 bp in length
27023 27122: gap of 100 bp
27123 27825: contig of 703 bp in length
27826 27925: gap of 100 bp
27926 28632: contig of 707 bp in length
28633 28732: gap of 100 bp
28733 29428: contig of 696 bp in length
29429 29528: gap of 100 bp
29529 30222: contig of 694 bp in length
30223 30322: gap of 100 bp
30323 31013: contig of 691 bp in length
31014 31113: gap of 100 bp
31114 31806: contig of 693 bp in length
31807 31906: gap of 100 bp
31907 32605: contig of 699 bp in length
32606 32705: gap of 100 bp
32706 33394: contig of 689 bp in length
33395 33494: gap of 100 bp
33495 34187: contig of 693 bp in length
34188 34287: gap of 100 bp
34288 34989: contig of 702 bp in length
34990 35089: gap of 100 bp
35090 35798: contig of 709 bp in length
35799 35898: gap of 100 bp
35899 36597: contig of 699 bp in length
36598 37398: contig of 701 bp in length
37399 37498: gap of 100 bp
37499 38207: contig of 709 bp in length
38208 38307: gap of 100 bp
38308 39006: contig of 699 bp in length
39007 39106: gap of 100 bp
39107 39803: contig of 697 bp in length
39804 39903: gap of 100 bp
39904 40623: contig of 720 bp in length
40624 40723: gap of 100 bp
40724 41415: contig of 692 bp in length
41416 41515: gap of 100 bp
41516 42212: contig of 697 bp in length

```

```

Query Match      73.9%: score 17; DB 2; length 63669;
Best local Similarity 100.0%: Pred. No. 17;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2 GGGCAATCTTACTTTC 18
Db      1341 GGGCAATCTTACTTTC 1325

```

```

42213 42312: gap of 100 bp
42313 43005: contig of 693 bp in length
43006 43105: gap of 100 bp
43106 43792: contig of 687 bp in length
43793 43892: gap of 100 bp
43893 44601: contig of 709 bp in length
44602 44701: gap of 100 bp
44702 45398: contig of 697 bp in length
45399 45498: gap of 100 bp
45499 46189: contig of 691 bp in length
46190 46289: gap of 100 bp
46290 46980: contig of 691 bp in length
46981 47080: gap of 100 bp
47081 47780: contig of 700 bp in length
47781 47880: gap of 100 bp
47881 48582: contig of 702 bp in length
48583 48682: gap of 100 bp
48683 49369: contig of 687 bp in length
49370 49469: gap of 100 bp
49470 50159: contig of 690 bp in length
50160 50259: gap of 100 bp
50260 50958: contig of 699 bp in length
50959 51058: gap of 100 bp
51059 51765: contig of 707 bp in length
51766 51865: gap of 100 bp
51866 52568: contig of 703 bp in length
52569 52668: gap of 100 bp
52669 53356: contig of 688 bp in length
53357 53456: gap of 100 bp
53457 54153: contig of 697 bp in length
54154 54253: gap of 100 bp
54254 54947: contig of 694 bp in length
54948 55047: gap of 100 bp

```

```

RESULT 6
LOCUS      AC124530
DEFINITION AC124530
ACCESSION  AC124530
VERSION    AC124530.2 GI:21414484
KEYWORDS   MUS MUSCULUS.
SOURCE     Mus musculus.
ORGANISM   Mus musculus.
REFERENCE  1 (bases 1 to 177182) DNA linear HTG 21-JUL-2002
AUTHORS   McPherson,J.D. and Waterston,R.H.
TITLE      The sequence of Mus musculus clone
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 177182)
AUTHORS   McPherson,J.D. and Waterston,R.H.
TITLE      Direct Submission
JOURNAL    Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
           Parkway, St. Louis, MO 63108, USA
REFERENCE  3 (bases 1 to 177182)
AUTHORS   McPherson,J.D. and Waterston,R.H.
TITLE      Direct Submission
JOURNAL    Submitted (21-JUL-2002) Genome Sequencing Center, 4444 Forest Park
           Parkway, St. Louis, MO 63108, USA
COMMENT    On Jul 21, 2002 this sequence version replaced gi:21426654.

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center

Center code: WUCSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.wustl.edu
Project information
Center project name: M.BA0286D07

Summary Statistics

Sequencing vector: M13, 08
Sequencing vector: plasmid: 1008
Chemistry: Dye-Primer ET: 0% of reads
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990319
Consensus quality: 174819 bases at least Q40
Consensus quality: 175383 bases at least Q30
Consensus quality: 175846 bases at least Q20
Insert size: 152000: agarose-fp
Insert size: 177561: sum-of-contigs
Quality coverage: 11.47 in Q20 bases; agarose-fp
Quality coverage: 9.32 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1180: contig of 1180 bp in length
* 1181 1280: gap of unknown length
* 1281 5117: contig of 3837 bp in length
* 5118 5217: gap of unknown length
* 5218 14456: contig of 9239 bp in length
* 14457 14556: gap of unknown length
* 14557 28704: contig of 14148 bp in length
* 28705 28804: gap of unknown length
* 28805 41022: contig of 12218 bp in length
* 41023 41123: gap of unknown length
* 41123 70598: contig of 29476 bp in length
* 70599 70698: gap of unknown length
* 70699 125895: contig of 55197 bp in length
* 125896 125995: gap of unknown length
* 125996 177182: contig of 51187 bp in length.

Location/Qualifiers

1. 177182

/organism="Mus musculus"

/db_xref="taxon:10090"

/chromosome="Ynk"

/clone="RP23-286D7"

1. 1180

/note="assembly_name:Contig17"

1281. 5117

/note="assembly_name:Contig18"

clone_end:SP6

vector_side:left

5218. 14456

/note="assembly_name:Contig19"

14557. 28704

/note="assembly_name:Contig10"

28805. 41022

/note="assembly_name:Contig11"

41123. 70598

/note="assembly_name:Contig12"

70699. 125895

/note="assembly_name:Contig13"

125996. 177182

/note="assembly_name:Contig14"

BASE COUNT 48045 a 39613 c 40407 g 48410 t 707 others

ORIGIN

Query Match 73.9% Score 17; DB 2; Length 177182;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0y 2 GCGCAATCTTACTTTC 18
Db 92751 GCGCAATCTTACTTTC 92767

RESULT 7

AC130947/c

LOCUS

DEFINITION

***, 86 unordered pieces.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 202953)

Muzny,D,Marle, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J,

Allen,C, Allen,H, Albrooks,S, Amin,A, Angiano,D,

Aoyakebechi,Y, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,

Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F,

Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,

Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E,

Cardenas,V, Carter,K, Cavazos,I, Caesar,H, Centler,A,

Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,

Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,

Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D,

Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K,

Draper,H, Dugan-Kocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K,

Egan,A, Escotlo,M, Eugene,C, Evans,C,A, Falls,T, Fan,G,

Fernandez,S, Finley,M, Flanagan,N, Forbes,L, Foster,M, Foster,P,

Fraser,C,M, Gabisi,A, Gant,R, Garcia,A, Garner,T, Garza,M,

Cebregregeis,E, Geer,K, Gill,R, Grady,M, Guerra,M, Guevara,M,

Gunaratne,P, Haaland,W, Hamli,C, Hamilton,C, Hamilton,K,

Harvey,Y, Havali,P, Hawes,A, Henderson,N, Hernandez,J,

Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hoque,M,

Hollins,B, Howell,S, Huylk,S, Hume,J, Idlebird,D, Jackson,A,

Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,K, Jolivet,A,

Karpachy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C,

Kovis,C, Kraft,C, Li, Lebow,H, Levan,J, Lewis,J, Li,Z, Liu,J,

Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J,

Lorenschew,L, Louisedge,H, Lozada,R,J, Lu,X, Ma,J,

Maheshwari,M, Mahindartine,M, Mahmoud,M, Mallory,K, Mangum,A,

Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E,

Manwiny,S, McLeod,M, McNeill,T, Meeneu,E, Milosavljevic,A,

Miner,G, Minja,E, Montemayor,J, Moore,S, Moran,M, Morris,K,

Morris,S, Mundasa,M, Murphy,M, Nair,L, Nankervis,C, Neal,D,

Newlon,N, Nguyen,N, Norris,S, Nwokediemo,O, Okwionu,G,

Olarinunsaogun,A, Pal,S, Parks,K, Pasternak,S, Paul,H,

Perez,A, Perez,L, Pfankuch,C, Plopper,F, Polindexter,A,

Popovic,D, Primus,E, Pu,L, Puzos,M, Quiroz,J, Rachlin,E,

Reeves,K, Regier,M,A, Reigh,K, Reilly,B, Reilly,M, Ren,Y,

Reuter,M, Richards,S, Riggs,F, Rives,C, Rodkey,T, Rojas,A,

Rose,M, Rose,R, Ruiz,S,J, Sanders,W, Savery,G, Scherer,S,

Scott,G, Shatsman,S, Shen,H, Shetty,J, Shvartsbeyn,A,

Sisson,I, Sitter,C,D, Smajs,D, Steele,A, Sodergren,E,

Song,X,Z, Sorrelle,R, Sosa,J, Steidle,M, Strong,K, Sutton,A,

Svatek,A, Tabor,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S,

Tingey,A, Trejos,Z, Usmani,K, Valas,R, Vera,V, Villasana,D,

Waldron,L, Walker,B, Wang,Q, Wang,S, Warren,J,

Warren,H, Wei,X, White,F, Williams,G, Willson,R, Wleczek,R,

Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S,

Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X,

Zhao,S, Dunn,D, von Niederhausern,A, Weiss,R,A,

Holt,R,A, Smith,H,O, Weinstock,G, and Gibbs,R,A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

2 (bases 1 to 202953)
Rat Genome Sequencing Consortium.
Direct Submission

JOURNAL
Submitted (15-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center Project name: KCBA

Center Clone name: CH230-199D23

Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap: version 0.990329

Consensus quality: 101023 bases at least Q40

Consensus quality: 113373 bases at least Q30

Consensus quality: 122379 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

consists of 86 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

```

1 1037: contig of 1037 bp in length
1038 1137: gap of unknown length
1138 2652: contig of 1515 bp in length
2653 2752: gap of unknown length
2753 4313: contig of 1561 bp in length
4314 4413: gap of unknown length
4414 5911: contig of 1498 bp in length
5912 6011: gap of unknown length
6012 7041: contig of 1030 bp in length
7042 7141: gap of unknown length
7142 8516: contig of 1375 bp in length
8517 8616: gap of unknown length
8617 10009: contig of 1393 bp in length
10010 10109: gap of unknown length
10110 11247: contig of 1138 bp in length
11248 11347: gap of unknown length
11348 12559: contig of 1212 bp in length
12560 12659: gap of unknown length
12660 14359: contig of 1700 bp in length
14360 14459: gap of unknown length
14460 15563: contig of 1104 bp in length
15564 15663: gap of unknown length
15664 17284: contig of 1621 bp in length
17285 17384: gap of unknown length
17385 18985: contig of 1601 bp in length
18986 19085: gap of unknown length
19086 20583: contig of 1498 bp in length
20584 22246: gap of unknown length
22247 22346: gap of unknown length
22347 23478: contig of 1132 bp in length
23479 23578: gap of unknown length
23579 25234: contig of 1656 bp in length
25235 25334: gap of unknown length
25335 26448: contig of 1114 bp in length
26449 28198: contig of 1650 bp in length
28199 28298: gap of unknown length
28299 29636: contig of 1338 bp in length
29637 29736: gap of unknown length
29737 31406: contig of 1670 bp in length
31407 31506: gap of unknown length
31507 32893: contig of 1387 bp in length
32894 32993: gap of unknown length

```

```

32994 34620: contig of 1627 bp in length
34621 34720: gap of unknown length
34721 35845: contig of 1125 bp in length
35846 35945: gap of unknown length
35946 37064: contig of 1119 bp in length
37065 37164: gap of unknown length
37165 38342: contig of 1178 bp in length
38343 38442: gap of unknown length
38443 40316: contig of 1874 bp in length
40317 40416: gap of unknown length
40417 41445: contig of 1029 bp in length
41446 41545: gap of unknown length
41546 42590: contig of 1045 bp in length
42591 42690: gap of unknown length
42691 44395: contig of 1705 bp in length
44396 44495: gap of unknown length
44496 45529: contig of 1034 bp in length
45530 45629: gap of unknown length
45630 47298: contig of 1669 bp in length
47299 47398: gap of unknown length
47399 48967: contig of 1569 bp in length
48968 49067: gap of unknown length
49068 50683: contig of 1616 bp in length
50684 50783: gap of unknown length
50784 52637: contig of 1854 bp in length
52638 52737: gap of unknown length
52738 55567: contig of 2830 bp in length
55568 55667: gap of unknown length
55668 57025: contig of 1358 bp in length
57026 57125: gap of unknown length
57126 58785: contig of 1660 bp in length
58786 58885: gap of unknown length
58886 60111: contig of 1226 bp in length
60112 60211: gap of unknown length
60212 61411: contig of 1200 bp in length
61412 61511: gap of unknown length
61512 62960: contig of 1449 bp in length
62961 63060: gap of unknown length
63061 65722: contig of 2662 bp in length
65723 65822: gap of unknown length
65823 67931: contig of 2109 bp in length
67932 68031: gap of unknown length
68032 69917: contig of 1886 bp in length
69918 70017: gap of unknown length
70018 71053: contig of 1036 bp in length
71054 71154: gap of unknown length
71155 72720: contig of 1567 bp in length
72721 72820: gap of unknown length
72821 74649: contig of 1849 bp in length
74650 74769: gap of unknown length
74770 76128: contig of 1459 bp in length
76129 76228: gap of unknown length
76229 77839: contig of 1611 bp in length
77840 77939: gap of unknown length
77940 80061: contig of 2122 bp in length
80062 80161: gap of unknown length
80162 81720: contig of 1559 bp in length
81721 81820: gap of unknown length
81821 83833: contig of 2013 bp in length
83834 83933: gap of unknown length
83934 85964: contig of 2031 bp in length
85965 86064: gap of unknown length

```

Query Match

Best Local Similarity 73.9%: Score 17; DB 2: Length 202953;

Matches 17: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGGCAAACTTACTTT 17

Db 191355 GGGGCAAACTTACTTT 191339

RESULT 8

AC122351/c 218369 bp DNA linear HTG 20-JUN-2002
 LOCUS AC122351
 DEFINITION Mus musculus chromosome UNK clone RP23-388H1, WORKING DRAFT
 SEQUENCE, 18 unordered pieces.
 AC122351
 AC122351.3 GI:21490497
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 218369)
 McPherson, J.D. and Waterston, R.H.
 The sequence of Mus musculus clone
 Unpublished
 2 (bases 1 to 218369)
 McPherson, J.D. and Waterston, R.H.
 Direct Submission
 Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 3 (bases 1 to 218369)
 McPherson, J.D. and Waterston, R.H.
 Direct Submission
 Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 On Jun 20, 2002 this sequence version replaced gi:21427982.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc/index.shtml>
 Contact: submissions@watson.wustl.edu
 Project Information
 Center project name: M_BA0388H01

----- Summary Statistics -----
 Sequencing vector: M13; 0x
 Sequencing vector: plasmid; 100x
 Chemistry: Dye-primer ET; 0% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 211350 bases at least Q40
 Consensus quality: 212482 bases at least Q30
 Consensus quality: 213333 bases at least Q20
 Insert size: 187000; agarose-fp
 Insert size: 216051; sum-of-ctrls
 Quality coverage: 13.42 in Q20 bases; agarose-fp
 Quality coverage: 9.60 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 18 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1112: contig of 1112 bp in length
 * 1113 1212: gap of unknown length
 * 1213 1830: contig of 618 bp in length
 * 1831 1930: gap of unknown length
 * 1931 15826: contig of 13896 bp in length
 * 15827 15925: gap of unknown length
 * 15927 17200: contig of 1274 bp in length
 * 17201 17301: gap of unknown length
 * 17301 18595: contig of 1295 bp in length
 * 18596 18695: gap of unknown length
 * 18696 21058: contig of 2363 bp in length
 * 21059 21159: gap of unknown length
 * 21159 22255: contig of 1097 bp in length
 * 22256 22356: gap of unknown length
 * 22356 24381: contig of 2026 bp in length
 * 24382 24481: gap of unknown length

FEATURES
 SOURCE

24482 28380: contig of 3899 bp in length
 * 28381 28480: gap of unknown length
 * 28481 30425: contig of 1945 bp in length
 * 30426 30525: gap of unknown length
 * 30526 37261: contig of 6736 bp in length
 * 37262 37362: gap of unknown length
 * 37363 48384: contig of 11023 bp in length
 * 48385 48485: gap of unknown length
 * 48486 64084: contig of 15600 bp in length
 * 64085 64184: gap of unknown length
 * 64185 85804: contig of 21620 bp in length
 * 85805 85904: gap of unknown length
 * 85905 115049: contig of 29145 bp in length
 * 115050 115149: gap of unknown length
 * 115150 141451: contig of 26302 bp in length
 * 141452 141551: gap of unknown length
 * 141552 165297: contig of 23746 bp in length
 * 165298 165398: gap of unknown length
 * 165399 218369: contig of 52972 bp in length.

location/qualifiers
 1. 218369
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="UNK"
 /clone="RP23-388H1"
 1. 1112
 /note="assembly_name:Contig12"
 1213. 1830
 /note="assembly_name:Contig14"
 1931. 15826
 /note="assembly_name:Contig11
 clone_end:17
 vector_size:17200
 15927. 17200
 /note="assembly_name:Contig23"
 17301. 18595
 /note="assembly_name:Contig24"
 18696. 21058
 /note="assembly_name:Contig25"
 21159. 22255
 /note="assembly_name:Contig26"
 22356. 24381
 /note="assembly_name:Contig27"
 24482. 28380
 /note="assembly_name:Contig28"
 28481. 30425
 /note="assembly_name:Contig29"
 30526. 37261
 /note="assembly_name:Contig30"
 37363. 48384
 /note="assembly_name:Contig32"
 48485. 64084
 /note="assembly_name:Contig33"
 64185. 85804
 /note="assembly_name:Contig34"
 85905. 115049
 /note="assembly_name:Contig35"
 115150. 141451
 /note="assembly_name:Contig36"
 141552. 165297
 /note="assembly_name:Contig37"
 165399. 218369
 /note="assembly_name:Contig38"

BASE COUNT 58332 a 49454 c 48698 g 60173 t 1712 others

ORIGIN

Query Match 73.9% Score 17; DB 2; Length 218369;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGCCAAATCTACTTTC 18
 |||
 Db 210282 GGCCAAATCTACTTTC 210266

RESULT 9
AC124772
LOCUS
DEFINITION
AC124772
AC124772
AC124772.1 GI:21428045
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

235301 bp DNA linear HTG 23-JUN-2002
Mus musculus chromosome UNK clone RP23-47P3, WORKING DRAFT
SEQUENCE, 19 unordered pieces.
AC124772
AC124772.1 GI:21428045
HTG: HTGS_PHASE1; HTGS_DRAFT.
house mouse.
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 235301)
McPherson, J.D. and Waterston, R.H.
The sequence of Mus musculus clone
unpublished
2 (bases 1 to 235301)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (15-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 235301)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (23-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M.BA0047P03

Summary Statistics
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 224743 bases at least Q40
Consensus quality: 226669 bases at least Q30
Consensus quality: 228045 bases at least Q20
Insert size: 213000; agarose-fp
Insert size: 234906; sum-of-contigs
Quality coverage: 8.20 in Q20 bases; agarose-fp
Quality coverage: 6.79 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1168: contig of 1168 bp in length
* 1169 1268: gap of unknown length
* 1269 2715: contig of 1447 bp in length
* 2716 2815: gap of unknown length
* 2816 4074: contig of 1258 bp in length
* 4075 4175: gap of unknown length
* 4175 5951: contig of 1777 bp in length
* 5952 6052: gap of unknown length
* 6052 8674: contig of 2623 bp in length
* 8675 8775: gap of unknown length
* 8775 12684: contig of 3910 bp in length
* 12685 12785: gap of unknown length
* 12785 15705: contig of 2920 bp in length
* 15705 15804: gap of unknown length

15805 19078: contig of 3274 bp in length
* 19079 19178: gap of unknown length
* 19179 25772: contig of 6594 bp in length
* 25773 25873: gap of unknown length
* 25873 34845: contig of 8972 bp in length
* 34845 34945: gap of unknown length
* 34945 46562: contig of 11618 bp in length
* 46563 46663: gap of unknown length
* 46663 64121: contig of 17458 bp in length
* 64121 64221: gap of unknown length
* 64221 87905: contig of 23664 bp in length
* 87905 88004: gap of unknown length
* 88004 107644: contig of 19640 bp in length
* 107645 107745: gap of unknown length
* 107745 126308: contig of 18564 bp in length
* 126309 126409: gap of unknown length
* 126409 145549: contig of 19141 bp in length
* 145550 145650: gap of unknown length
* 145650 16095: contig of 20445 bp in length
* 16095 16194: gap of unknown length
* 16195 192987: contig of 26793 bp in length
* 192988 235301: contig of 42214 bp in length.
193088

FEATURES
Source
1..235301
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-47P3"
1..1168
/note="assembly_name:Contig3"
1269..2715
/note="assembly_name:Contig15"
2816..4074
/note="assembly_name:Contig16"
4175..5951
/note="assembly_name:Contig17"
6052..8674
/note="assembly_name:Contig18"
8775..12684
/note="assembly_name:Contig19"
12785..15704
/note="assembly_name:Contig20"
clone_end:Spb
vector_side:left
15805..19078
/note="assembly_name:Contig21"
19179..25772
/note="assembly_name:Contig22"
25873..34844
/note="assembly_name:Contig23"
34945..46562
/note="assembly_name:Contig24"
46663..64120
/note="assembly_name:Contig25"
64221..87904
/note="assembly_name:Contig26"
88005..107644
/note="assembly_name:Contig27"
107745..126308
/note="assembly_name:Contig28"
126409..145549
/note="assembly_name:Contig29"
145650..16094
/note="assembly_name:Contig30"
16195..192987
/note="assembly_name:Contig31"
193088..235301
/note="assembly_name:Contig32"
BASE COUNT 63617 a 52537 c 53697 g 63522 t 1928 others
Query Match 73.9%; Score 17; DB 2; Length 235301;

Best Local Similarity 100.0%; Pred. No. 14;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GGGCAATCTTACTTTC 18
|||||
Db 101622 GGGCAATCTTACTTTC 101638

RESULT 10
AC124601 239619 bp DNA linear HTG 08-AUG-2002
LOCUS Mus musculus chromosome UNK clone RP23-27M3, WORKING DRAFT
DEFINITION
SEQUENCE 7 unordered pieces.
AC124601
AC124601.2 GI:22138700
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 239619)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 239619)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 239619)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (08-AUG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Aug 8, 2002 this sequence version replaced gi:21426743.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M_BA0027M03
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 237244 bases at least Q40
Consensus quality: 237642 bases at least Q40
Consensus quality: 237926 bases at least Q20
Insert size: 213000; agarose-fp
Insert size: 239019; sum-of-contigs
Quality coverage: 11.66 in Q20 bases; agarose-fp
Quality coverage: 9.65 in Q20 bases; sum-of-contigs
----- NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1687: contig of 1687 bp in length
* 1688 1787: gap of unknown length
* 1788 7817: contig of 6030 bp in length
* 7818 7917: gap of unknown length
* 7918 96227: contig of 88310 bp in length
* 96228 96328: gap of unknown length
* 96329 112836: contig of 16509 bp in length

* 112837 112936: gap of unknown length
* 112937 140009: contig of 27073 bp in length
* 140010 140109: gap of unknown length
* 140110 166594: contig of 26485 bp in length
* 166595 166695: gap of unknown length
* 166696 239619: contig of 72925 bp in length.
Location/Qualifiers
1..239619
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-27M3"
1..1687
/note="assembly_name:Contig23"
1788..7817
/note="assembly_name:Contig24
clone_end:5pb
vector_side:left"
7918..96227
/note="assembly_name:Contig29"
96328..112836
/note="assembly_name:Contig25"
112937..140009
/note="assembly_name:Contig26"
140110..166594
/note="assembly_name:Contig27"
166695..239619
/note="assembly_name:Contig28"

BASE COUNT 65204 a 52622 c 54054 g 67126 t 613 others
ORIGIN

Query Match 73.9%; Score 17; DB 2; Length 239619;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGGCAATCTTACTTTC 18
|||||
Db 127590 GGGCAATCTTACTTTC 127606

RESULT 11
AX183106/c
LOCUS AX183106 700 bp DNA linear PAT 06-AUG-2001
DEFINITION Sequence 859 from Patent WO0142511.
ACCESSION AX183106
VERSION AX183106.1 GI:15134407
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 700)
AUTHORS Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae; Mus.
TITLE Inbred-related polymorphisms
JOURNAL Patent: WO 0142511-A 859 14-JUN-2001;
WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH (US) : ELLIPSIS
Biotherapeutics Corporation (CA)
FEATURES
source 1..700
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 274 a 88 c 115 g 219 t 4 others
ORIGIN

Query Match 69.6%; Score 16; DB 6; Length 700;
Best Local Similarity 100.0%; Pred. No. 1; 2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GGGCAATCTTACTTTC 17
|||||
Db 385 GGGCAATCTTACTTTC 370

```

RESULT 12
AF086808      2238 bp  mRNA  linear  MAM 06-OCT-1998
LOCUS         AF086808
DEFINITION    Bos taurus fertillin beta (ADAM 2) mRNA, complete cds.
ACCESSION     AF086808
VERSION       AF086808.1 GI:3661511
KEYWORDS
SOURCE        Bos taurus.
ORGANISM      Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE     1 (bases 1 to 2238)
AUTHORS      Waters,S.I. and White,J.M.
TITLE        Biochemical and molecular characterization of bovine fertillin alpha
              and beta (ADAM 1 and ADAM 2): a candidate sperm-egg binding/fusion
              complex
JOURNAL       Biol. Reprod. 56 (5), 1245-1254 (1997)
MEDLINE       97304362
PUBMED        9160725
REFERENCE     2 (bases 1 to 2238)
AUTHORS      Waters,S.I. and White,J.M.
TITLE        Direct Submission
JOURNAL       Submitted (25-AUG-1998) Department of Cell Biology, University of
              Virginia, Health Sciences Center, Box 439, Charlottesville, VA
              22908, USA

FEATURES
source        Location/Qualifiers
              1..2238
                /organism="Bos taurus"
                /db_xref="taxon:9913"
                /tissue_type="testis"
              1..2238
                /gene="ADAM 2"
              1..2238
                /gene="ADAM 2"
                /codon_start=1
                /product="fertillin beta"
                /protein_id="AAC62753.1"
                /db_xref="GI:3661512"
              /translation="MCLFLFLGLGLTQTDNSENRLRVQTPVEKIRSTSGGVETH
              VSYILLIEGKTVTNLMOKAFIPHNHRYVYSGTSGSMKLEHEFONFCYQYIEGYP
              NSMALISTGTGLRGLDFENVSGLIEPLPSIGFEFMYOIKRDRSSSVYERHLEL
              REPKYKIONVPLPDPFSOYIEHIVYERKDLVHMKADTVVIOKIFQLGLTNAIFTS
              LNTVITYLSIELMIDENKIPVIGDANELLHRVKKRSTLVLRPDMALVYRKRSN
              YIGATFGQKMDCKHYGGVALHSTISLESLAVIIAQLLSIMGIPYDINCHGPD
              VCIMNPAVHSSGVKLFNSCSEDEFLRETSKPSOCLQNPRLDPTPKSAVCGNGKVE
              EGEGDCGCKKACDALPDTCCVADTCRFQPGSACDTGLCCSCAFIPKGIICGSTDE
              CDLHEYCGSSAACQEDVYVODGHCPCGNOMLCTIGICVDGICKECEDIEGEGTSVA
              ECFORLNSMNDISGNCVPTGCTPTCTSPNRCGLICFYDKREVISVSNASVMSYNSI
              NCKTICLHAYEGNDEBGMVWKDTCVCESTCKQKQCVDSSTLNYDCNPKCNMG
              VCNNKHCCHNFSYLPNCEHISAPGMEGSDISGNFPPSEPTGAPAFDVCFTPLAE
              SRYLENVKSKPTNMFPLEIFPEFIIICVLATLVKVFQRRKKRMTEDYTSDEQLESE
              SEPKD"

BASE COUNT   682 a 397 c 495 g 664 t
ORIGIN
Query Match 69.6%; Score 16; DB 4; Length 2238;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 GGCAGATCTTACTTTC 18
    |||
Db 1801 GGCAGATCTTACTTTC 1786

RESULT 13
BC030525      2558 bp  mRNA  linear  PRI 20-MAY-2002
LOCUS         BC030525
DEFINITION    Homo sapiens, similar to LOC201361, clone MGC:40524 IMAGE:5207811,
              mRNA, complete cds.
ACCESSION     BC030525
VERSION       BC030525.1 GI:20987814
KEYWORDS      MGC.

```

```

SOURCE        Homo sapiens.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE     1 (bases 1 to 2558)
AUTHORS      Strausberg,R.
TITLE        Direct Submission
JOURNAL       Submitted (07-MAY-2002) National Institutes of Health, Mammalian
              Gene Collection (MGC), Cancer Genomics Office, National Cancer
              Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
              USA
REMARK        NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT       Contact: MGC help desk
              Email: mgc@nci.nih.gov
              Tissue Procurement: Life Technologies, Inc.
              CDNA Library Preparation: Life Technologies, Inc.
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILM)
              DNA Sequencing by: National Institutes of Health Intramural
              Sequencing Center (NISC),
              Gaithersburg, Maryland;
              Web site: http://www.nisc.nih.gov/
              Contact: nisc.mgc@nih.gov
              Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
              Blakesley,K.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
              Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
              Hansen,N., Ho,S.-L., Karlins,E., Latic,P., Legaspi,R., Maduro,Q.L.,
              Masello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J.,
              Pearson,R., Stantipop,S., Thomas,P.J., Touchman,J.W., Tsurgoun,C.,
              Vogt,J.L., Walker,M.A., Wetherly,K.D., Wiggins,L., Young,A.,
              Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILM at: http://image.llnl.gov
Series: IRAC Plate: 64 Row: 1 Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

FEATURES
source        Location/Qualifiers
              1..2558
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="MGC:40524 IMAGE:5207811"
                /tissue_type="Lung, spleen, fetal, pooled"
                /clone_1fb="NIH_MGC_122"
                /lab_host="DH10B"
                /note="Vector: pcMV-SPOrte"
                /codon_start=1
                /product="Similar to LOC201361"
                /protein_id="AAH30525.1"
                /db_xref="GI:20987815"
                /translation="MRLASQRRRRRGKRRKKRRKKRRKKRRKKRRKKRRKKRR
                RRRGRRRRMQQ"

BASE COUNT   961 a 390 c 536 g 671 t
ORIGIN
Query Match 69.6%; Score 16; DB 9; Length 2558;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGCAGATCTTACTTT 17
    |||
Db 1819 GGCAGATCTTACTTT 1804

RESULT 14
AC004649      24102 bp  DNA  linear  PRI 01-MAY-1998
LOCUS         AC004649
DEFINITION    Homo sapiens chromosome 5, p1 clone 603E12 (LBNI H26), complete
              sequence.
ACCESSION     AC004649
VERSION       AC004649.1 GI:3097835
KEYWORDS      HTG.

```

SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1. (bases 1 to 24102)
AUTHORS Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K. S., Gunning, K. M., Davis, C. A., Kadner, K., Miquel, T., Plick, S., Pollard, M., Rojaski, H., Subramanian, S. and Martin, C. H.
TITLE Sequencing of human chromosome 5
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 24102)
TITLE Large Scale Sequence Analysis and Annotation with the Sequence Comparison Analysis (SCAN) System
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 24102)
AUTHORS Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K. S., Gunning, K. M., Davis, C. A., Kadner, K., Miquel, T., Plick, S., Pollard, M., Rojaski, H., Subramanian, S. and Martin, C. H.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-1998) Human Genome Center, DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, MS 74-157, Berkeley, CA 94720, U.S.A.
COMMENT Sequence submitted by:
DOE Joint Genome Institute.
FEATURES
source
1. 24102
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q"
/clone="603E12"
/note="LBNL H26"
60. 417
/rpt_family="Alu"
repeat_region
1739. 2075
/rpt_family="Alu"
repeat_region
2048. 2073
/note="(A)26"
/rpt_type=tandem
/rpt_unit=A
complement(2187. 2520)
repeat_region
/rpt_family="Alu"
complement(3077. 3228)
repeat_region
/rpt_family="MSTB"
complement(3307. 3433)
repeat_region
/rpt_family="MSTC"
complement(4377. 4447)
repeat_region
/rpt_family="L1"
6154. 6318
/rpt_family="Alu"
repeat_region
10228. 10476
/note="100% identity A00906"
join(11117. 11800,11901. 12034)
repeat_region
/note="protein similarities to LIM protein RIL"
11940. 12032
complement(13226. 13520)
STS
/db_xref="dbSTS:252623"
13344. 13371
/note="(GT)14"
repeat_region
/rpt_type=tandem
/rpt_unit=GT
16236. 16301
/note="(TTC)22"
repeat_region
/rpt_type=tandem
/rpt_unit=TTTC
join(16641. 16884,17101. 17373,18904. 18962,22891. 23040)
misc_feature
/note="multiple regions of similarity to 272522"
misc_feature
16990. 16774
/note="82% identity Z95114"
misc_feature
16753. 17018
/note="82% identity ESM6275b"

misc_feature /db_xref="dbEST:AA354486"
complement(16807. 17033)
misc_feature /note="82% identity H59082"
complement(17020. 17100)
misc_feature /note="GRAIL 2 excellent exon, frame 0"
17101. 17373
misc_feature /note="81% identity Z75522"
join(17114. 17373,17580. 17705,18001. 18260,19972. 20031, 21599. 21700,22501. 22639)
repeat_region /note="multiple regions of similarity to 284474"
18431. 18707
repeat_region /rpt_family="Alu"
complement(20880. 21211)
repeat_region /rpt_family="Alu"
complement(21188. 21246)
misc_feature /note="GRAIL 2 excellent exon, frame 1"
22074. 22159
misc_feature /note="81% identity yH96d01.r1"
/db_xref="dbEST:R68365"
join(22180. 22600,22715. 23102)
prim_transcript /standard_name="zh69e03"
/note="98% identity W89188 and W8931"
22988. 23177
misc_feature /note="81% identity z13b02.s1"
/db_xref="dbEST:AA775691"
complement(23203. 23500)
misc_feature /note="81% identity y119906.r1"
/db_xref="dbEST:R64030"
complement(23586. 23890)
repeat_region /rpt_family="Alu"
BASE COUNT 6162 a 5923 c 5660 g 6357 t
ORIGIN
Query Match 69.6% Score 16: DB 9: Length 24102:
Best Local Similarity 100.0% Pred. No. 70:
Matches 16: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
QY 2 GGCGCAATCTTACTTT 17
IIIIIIIIIIIIIIIIII
Db 3795 GGCGCAATCTTACTTT 3810
RESULT 15
CEC05C9 33734 bp DNA linear INV 13-AUG-2002
LOCUS Cenorhabditis elegans cosmid C05C9, complete sequence.
DEFINITION 768000
ACCESSION 768000
VERSION 768000.1 GI:1070038
KEYWORDS HTG; NMU1.
SOURCE Cenorhabdilis elegans.
OK/ANISM Cenorhabdilis elegans.
REFERENCE 1
AUTHORS none.
TITLE Genome sequence of the nematode C. elegans: a platform for
JOURNAL Investigating biology. The C. elegans Sequencing Consortium
MEDLINE Science 282 (5396), 2012-2018 (1998)
REMARK 99069613
REFERENCE The C. elegans Sequencing Consortium.
AUTHORS 2 (bases 1 to 33734)
TITLE Kershaw, J.K.
JOURNAL Direct Submission
SUBMITTED (21-NOV-1995) Nematode Sequencing Project, Sanger
Institute, Hinxton, Cambridge CB10 1SA, England and Department of
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
jes@wustl.ac.uk or w@nematode.wustl.edu
Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a

dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is not the entire insert of clone C05C9. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone C05C9 is at 1 in this sequence. The true right end of clone C05C9 is at 7767 in sequence 250739.

The true left end of clone F13D2 is at 33731 in this sequence. The true right end of clone R08B4 is at 15083 in this sequence. The start of this sequence (1..110) overlaps with the end of sequence 268008.

The end of this sequence (33731..33734) overlaps with the start of sequence 250739.

For a graphical representation of this sequence and its analysis see: <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=C05C9>

FEATURES

source

1..33734

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="X"

/clone="C05C9"

complement(join(9704..9904,10252..10464,10516..10764,10912..11010,11134..11172,11218..11370,12454..12528,12580..12648,12880..13325,13597..13659,13708..14466,14743..15231,15277..15559,16139..16223,16264..16396,16439..16553,18349..18491,19079..19142))

/gene="C05C9.3"
complement(join(9704..9904,10252..10464,10516..10764,10912..11010,11134..11172,11218..11370,12454..12528,12580..12648,12880..13325,13597..13659,13708..14466,14743..15231,15277..15559,16139..16223,16264..16396,16439..16553,18349..18491,19079..19142))

/note="Similarity to Yeast NMI protein (PIR Acc. NO. S19052)."

/codon_start=-1

/protein_id="CA91971.1"

/db_xref="GI:3874019"

/db_xref="SPTREMBL:O17656"

/translation="MEDLESTSDNEENHRLTLXDKTQSKWYLNKAKTIKSEPOIEK

LGRLLFLDDETHDKLIDYISIVYLNKLNELPVDTGCPYNTMTPECGTSTSHOPS

MISADDRKFSVIAKIGVFDHPQEFMDIIPFETAGDETHVIASTIFRYARKFSP

SISGVIRIKTLRLNNGRPSRKAIVLLVHPQIIPICIGKHKFKMDYVIMVISM

STLSTICCKNNMGIFMHIQENHLYKSPGFLMGCPHHLILLVACANFKIMVKNP

KIRKAERKTLPIKLELLETLLKPKPKPCAEEDVYIRGKAQVNSNIEVYIDH

DDIGINQEDKTKROSADDEECDLNVAFCNSNTEIKKAEDELISHLESSONS

IHNDRKKNDYVSELETETTOROLEGCTSPENKMERITVNSPTDILSOTIM

KHPQNLPEVETSSSEITKYGQIKSPFQSENGIKVIDEGLKSPSLNVEVSEDC

WSPFQRTISCDGVSSESTRNNKAPQVSPKQAKKQVISENFGESEQSDC

ETSNKSKSDENMLDVOEDNTSDNEISFPDEAPENRVTISTDLACDILLATIT

DSIHLTKVETVPEPIELVSNKKEGTGKEPEANGINPESIKOSSEVSESM

CSPDKPDDAPKQSDSKQDNQSEIPREKILDMAGECIETKILKMEETINCDAD

FELQKAEVGLKHSDSRNQADKRGKMKKLOHLVKTRLADPELTNAQOTWLLNMC

SNMPTVGLLEFTEFNENKCDTADQVYMTLPSFATIDQENKIEERASIPKSN

LEPVETLAKNELTDEFOKSTGCOQOVOTTEKHAOTGDEETSHLSNVSGLPML

SPRVLFKDHNETPQNIIMPVHNOVRLTVOVNSSQVAMALSVPAHOOHDEPL

MHEELVHOQPLKROMHEKMOQVNSIDEKFMYHQNLOOLROYEDOMOLOKOV

QEOHDOIRHHQMLKQERONHNMHDIIEFHOOHLNKRYPHOOGEKRODGMQ

MONOLHNHONFQNSGOATOVHEOQONHNVRAOYEEVFSRPHDVRHIGOT

NYOONORHOSQOOVYKFMQPLRIQPVNOQVQPNPNRPOFYPNNHGHVQVL

HOOHMTVNNNHAGFTNPS"

join(19396..19498,19548..19607,19650..19705,23207..23408,

23462..23513,23558..23640,23705..23740,23791..23873)

/gene="C05C9.2"

join(19396..19498,19548..19607,19650..19705,23207..23408,

23462..23513,23558..23640,23705..23740,23791..23873)

/gene="C05C9.2"

join(19396..19498,19548..19607,19650..19705,23207..23408,

23462..23513,23558..23640,23705..23740,23791..23873)

/gene="C05C9.2"

join(19396..19498,19548..19607,19650..19705,23207..23408,

23462..23513,23558..23640,23705..23740,23791..23873)

/gene="C05C9.2"

join(19396..19498,19548..19607,19650..19705,23207..23408,

23462..23513,23558..23640,23705..23740,23791..23873)

/gene="C05C9.2"

join(19396..19498,19548..19607,19650..19705,23207..23408,

23462..23513,23558..23640,23705..23740,23791..23873)

/gene="C05C9.2"

join(19396..19498,19548..19607,19650..19705,23207..23408,

23462..23513,23558..23640,23705..23740,23791..23873)

/gene="C05C9.2"

join(19396..19498,19548..19607,19650..19705,23207..23408,

23462..23513,23558..23640,23705..23740,23791..23873)

/gene="C05C9.2"

join(19396..19498,19548..19607,19650..19705,23207..23408,

23462..23513,23558..23640,23705..23740,23791..23873)

gene

/note="cDNA EST EMBL:BJ113191 comes from this gene
cDNA EST EMBL:BJ141026 comes from this gene"
/codon_start=-1
/protein_id="CA91970.2"
/db_xref="GI:22265667"
/db_xref="SPTREMBL:O17655"
/translation="MYREMHHQCIQLYCEIWPCKREKREHPIHTTKLRKMTSE
SYDHOIPMTGSKSKTGVEFEIMEMAKRQSONPTEVPLSLPEFQSLTIGSKRK
ROSTKADDTMIIFDNPVYDIKAFROGLADEDEKYSIGAPYIMDDDTLIDILROT
YHESASPIHYGTDETISSIDENYSTLIDKQANGSKRSKLLVPSLRSDAKYE
GSLIDL"

CDS

join(27311..27349,27403..27516,27570..27617,27752..28003,
28055..28126,28176..28250,28297..28528,28578..28674,
28822..29005,29050..29202,29819..30016)
/gene="C05C9.1"
join(27311..27349,27403..27516,27570..27617,27752..28003,
28055..28126,28176..28250,28297..28528,28578..28674,
28822..29005,29050..29202,29819..30016)
/note="contains similarity to Pfam domain: P001273 (4BP /
BPI / CERP family), Score=-75.6, E-value=0.00044, N=1
cDNA EST EMBL:M69420 comes from this gene"

source

1..33734

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="X"

/clone="C05C9"

complement(join(9704..9904,10252..10464,10516..10764,10912..11010,11134..11172,11218..11370,12454..12528,12580..12648,12880..13325,13597..13659,13708..14466,14743..15231,15277..15559,16139..16223,16264..16396,16439..16553,18349..18491,19079..19142))

gene

/note="Similarity to Yeast NMI protein (PIR Acc. NO. S19052)."
/codon_start=-1
/protein_id="CA91971.1"
/db_xref="GI:3874019"
/db_xref="SPTREMBL:O17656"
/translation="MEDLESTSDNEENHRLTLXDKTQSKWYLNKAKTIKSEPOIEK

gene

LGRLLFLDDETHDKLIDYISIVYLNKLNELPVDTGCPYNTMTPECGTSTSHOPS
MISADDRKFSVIAKIGVFDHPQEFMDIIPFETAGDETHVIASTIFRYARKFSP
SISGVIRIKTLRLNNGRPSRKAIVLLVHPQIIPICIGKHKFKMDYVIMVISM
STLSTICCKNNMGIFMHIQENHLYKSPGFLMGCPHHLILLVACANFKIMVKNP
KIRKAERKTLPIKLELLETLLKPKPKPCAEEDVYIRGKAQVNSNIEVYIDH
DDIGINQEDKTKROSADDEECDLNVAFCNSNTEIKKAEDELISHLESSONS
IHNDRKKNDYVSELETETTOROLEGCTSPENKMERITVNSPTDILSOTIM
KHPQNLPEVETSSSEITKYGQIKSPFQSENGIKVIDEGLKSPSLNVEVSEDC
WSPFQRTISCDGVSSESTRNNKAPQVSPKQAKKQVISENFGESEQSDC
ETSNKSKSDENMLDVOEDNTSDNEISFPDEAPENRVTISTDLACDILLATIT
DSIHLTKVETVPEPIELVSNKKEGTGKEPEANGINPESIKOSSEVSESM
CSPDKPDDAPKQSDSKQDNQSEIPREKILDMAGECIETKILKMEETINCDAD
FELQKAEVGLKHSDSRNQADKRGKMKKLOHLVKTRLADPELTNAQOTWLLNMC
SNMPTVGLLEFTEFNENKCDTADQVYMTLPSFATIDQENKIEERASIPKSN
LEPVETLAKNELTDEFOKSTGCOQOVOTTEKHAOTGDEETSHLSNVSGLPML
SPRVLFKDHNETPQNIIMPVHNOVRLTVOVNSSQVAMALSVPAHOOHDEPL
MHEELVHOQPLKROMHEKMOQVNSIDEKFMYHQNLOOLROYEDOMOLOKOV
QEOHDOIRHHQMLKQERONHNMHDIIEFHOOHLNKRYPHOOGEKRODGMQ
MONOLHNHONFQNSGOATOVHEOQONHNVRAOYEEVFSRPHDVRHIGOT
NYOONORHOSQOOVYKFMQPLRIQPVNOQVQPNPNRPOFYPNNHGHVQVL
HOOHMTVNNNHAGFTNPS"

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

lRNA

join(19396..19498,19548..19607,19650..19705,23207..23408,
23462..23513,23558..23640,23705..23740,23791..23873)

CDS

```
complement(32244..32312),complement(31998..32188),  
complement(31866..31953),complement(31439..31619),  
complement(31244..31390),complement(30713..30937),  
complement(30546..30666),complement(30409..30502))  
/gene="F13D2.1"  
join(complement(250739..1:7677..7748),  
complement(250739..1:6642..7048),  
complement(250739..1:6642..6818),  
complement(250739..1:4814..4972),  
complement(250739..1:3274..3441),  
complement(250739..1:2881..3045),  
complement(250739..1:2386..2497),  
complement(250739..1:2186..2311),  
complement(250739..1:2023..2098),  
complement(250739..1:1727..1816),  
complement(250739..1:1528..1675),  
complement(250739..1:1206..1477),  
complement(250739..1:1101..1158),  
complement(250739..1:932..1049),  
complement(250739..1:545..655),  
complement(250739..1:323..499),  
complement(250739..1:171..273),complement(250739..1:5..125),  
complement(33618..33734),complement(33419..33566),  
complement(33128..33222),complement(32359..32577),  
complement(32244..32312),complement(31998..32188),
```

Query Match

Best Local Similarity 69.6% Score 16: DB 3: Length 33734:
Matches 16: Conservative 0: Mismatches 0: Indels 0: Gaps 0:OY 3 GGCAATCTTACTTTC 18
|||||

Db 23921 GGCAATCTTACTTTC 23936

Search completed: January 11, 2003, 23:57:22
Job time : 3321 secs

```

; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 Unviersity Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/549,515
; FILING DATE: 27-OCT-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-522
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4051 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(65..2482, 2496..3110, 3115..3951)
US-08-549-515-1

```

```

Query Match          56.5%; Score 13; DB 3; Length 4051;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
Oy 5 CAAATCTTACTTT 17
   |||

```

```
Db 138 CAAATCTTACTTT 150

```

```

Search completed: January 11, 2003, 23:01:58
Job time : 70 secs

```

Db 516 AAATCTTACTTTC 504

RESULT 12
US-09-134-001C-1731/C
; Sequence 1731, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1731
; LENGTH: 2073
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1731

Query Match 56.5%; Score 13; DB 4; Length 2073;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 CAAATCTTACTTT 17
|||||Db 1595 CAAATCTTACTTT 1583
|||||

RESULT 13
US-08-549-515-2
; Sequence 2, Application US/08549515
; Patent No. 6054123
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; TITLE OF INVENTION: Haemophilus Influenzae
; TITLE OF INVENTION: Dimethylsulphoxide Reductase Enzyme
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/549,515
; FILING DATE: 27-OCT-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-522
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2418 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-08-549-515-2

Query Match 56.5%; Score 13; DB 3; Length 2418;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 CAAATCTTACTTT 17
|||||Db 74 CAAATCTTACTTT 86
|||||

RESULT 14
US-08-961-527-138/c
; Sequence 138, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 138:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-138

Query Match 56.5%; Score 13; DB 4; Length 3083;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 CAAATCTTACTTT 17
|||||Db 279 CAAATCTTACTTT 267
|||||

RESULT 15
US-08-549-515-1
; Sequence 1, Application US/08549515
; Patent No. 6054123
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; TITLE OF INVENTION: Haemophilus Influenzae
; TITLE OF INVENTION: Dimethylsulphoxide Reductase Enzyme
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:

ATTORNEY/AGENT INFORMATION:

NAME: Gimmil, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 167:

SEQUENCE CHARACTERISTICS:

LENGTH: 677 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-858-207A-167

Query Match

Best Local Similarity 100.0%; Score 13; DB 4; Length 677;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 AAATCTTACTTTC 18
|||||

Db 667 AAATCTTACTTTC 655

RESULT 9

US-09-175-658B-23/C

Sequence 23, Application US/09175658B

Patent No. 6372900

GENERAL INFORMATION:

APPLICANT: METALLINOS, DANIKA

APPLICANT: RINE, JASPER

APPLICANT: BOWLING, ANN

TITLE OF INVENTION: HORSE ENDOTHELIN-B RECEPTOR GENE AND GENE PRODUCTS

FILE REFERENCE: G08R-110

CURRENT APPLICATION NUMBER: US/09/175,658B

CURRENT FILING DATE: 1998-10-20

PRIOR APPLICATION NUMBER: 60/062,562

PRIOR FILING DATE: 1997-10-21

NUMBER OF SEQ ID NOS: 25

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 23

LENGTH: 752

TYPE: DNA

ORGANISM: Horse

FEATURE:

OTHER INFORMATION: Uncertain of the nucleotide sequence at positions

US-09-175-658B-23

Query Match

Best Local Similarity 100.0%; Score 13; DB 4; Length 752;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 TTACTTCGCTCC 23
|||||

Db 558 TTACTTCGCTCC 546

RESULT 10

US-09-134-001C-1728/C

Sequence 1728, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 1728

LENGTH: 765

TYPE: DNA

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-1728

Query Match

Best Local Similarity 100.0%; Score 13; DB 4; Length 765;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 AAATCTTACTTTC 18
|||||

Db 333 AAATCTTACTTTC 321

RESULT 11

US-08-793-035-3/C

Sequence 3, Application US/08794045

Patent No. 6011201

GENERAL INFORMATION:

APPLICANT: Slabas, Antoni R.

APPLICANT: White, Andrew

APPLICANT: Chase, Diane

APPLICANT: Elborough, Kelran

APPLICANT: Fentem, Phillip A.

TITLE OF INVENTION: B-Ketoadyl ACP Reductase Genes from

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold White & Burke

STREET: P.O. Box 4433

CITY: Houston

STATE: TX

COUNTRY: US

ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/793,045

FILING DATE: 28-JUL-1997

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9414622.2

FILING DATE: 20-JUL-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB95/01678

FILING DATE: 17-JUL-1995

ATTORNEY/AGENT INFORMATION:

NAME: Kammerer, Patricia A.

REGISTRATION NUMBER: 29,775

REFERENCE/DOCKET NUMBER: MOFT:132

TELECOMMUNICATION INFORMATION:

TELEPHONE: 713.787.1400

TELEFAX: 713.787.1440

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1299 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-793-035-3

Query Match

Best Local Similarity 100.0%; Score 13; DB 3; Length 1299;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 AAATCTTACTTTC 18
|||||

TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE: polynucleotide
US-08-365-981-1

Query Match 56.5%; Score 13; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 AATCTTACTTTG 19
|||||
DB 156 AATCTTACTTTG 144

RESULT 5
US-09-134-001C-96
Sequence 96, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 96
LENGTH: 339
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-96

Query Match 56.5%; Score 13; DB 4; Length 339;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 CAAATCTTACTTT 17
|||||
DB 85 CAAATCTTACTTT 97

RESULT 6
US-09-134-001C-131
Sequence 131, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 131
LENGTH: 339
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-131

Query Match 56.5%; Score 13; DB 4; Length 339;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 CAAATCTTACTTT 17

DB 85 CAAATCTTACTTT 97
|||||

RESULT 7
US-09-385-982-148/C
Sequence 148, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
PRODUCTS: 11
FILE REFERENCE: CCDA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 148
LENGTH: 614
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(614)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-148

Query Match 56.5%; Score 13; DB 4; Length 614;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 CAAATCTTACTTT 17
|||||
DB 304 CAAATCTTACTTT 292

RESULT 8
US-08-858-207A-167/C
Sequence 167, Application US/08858207A
Patent No. 6348328
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Slodola, Robert
TITLE OF INVENTION: NO. 6348328el Compounds
NUMBER OF SEQUENCES: 552
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,207A
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996

|||||
Db 21156 CTTACTTCCTTC 21169

RESULT 2

US-07-795-859B-10/C
Sequence 10, Application US/07795859B

Patent No. 5422262

GENERAL INFORMATION:

APPLICANT: Anderson, Stefan

APPLICANT: Russell, David W.

TITLE OF INVENTION: Steroid 5-Reductases

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White and Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: TX

COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII-DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/795,859B

FILING DATE: 18-NOV-1991

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Parker, David L.

REGISTRATION NUMBER: 32,165

REFERENCE/DOCKET NUMBER: UTSD:260/PAR

TELECOMMUNICATION INFORMATION:

TELEPHONE: (512) 320-7200

TELEFAX: (512) 474-7677

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 100 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-07-795-859B-10

Query Match 56.5%; Score 13; DB 1; Length 100;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 AAATCTTACTTC 18
|||||
Db 16 AAATCTTACTTC 4

RESULT 3

US-08-457-616-10/C
Sequence 10, Application US/08457616

Patent No. 5679521

GENERAL INFORMATION:

APPLICANT: Anderson, Stefan

APPLICANT: Russell, David W.

TITLE OF INVENTION: Steroid 5-Reductases

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White and Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: TX

COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII-DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/457,616

FILING DATE: 01-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/795,859

FILING DATE: 18-NOV-1991

ATTORNEY/AGENT INFORMATION:

NAME: Parker, David L.

REGISTRATION NUMBER: 32,165

REFERENCE/DOCKET NUMBER: UTSD:260/PAR

TELECOMMUNICATION INFORMATION:

TELEPHONE: (512) 320-7200

TELEFAX: (512) 474-7677

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 100 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-457-616-10

Query Match 56.5%; Score 13; DB 1; Length 100;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 AAATCTTACTTC 18
|||||
Db 16 AAATCTTACTTC 4

RESULT 4

US-08-365-981-1/C
Sequence 1, Application US/08365981

Patent No. 5583030

GENERAL INFORMATION:

APPLICANT: Robert DICKSON et al

TITLE OF INVENTION: METHOD FOR OBTAINING

TITLE OF INVENTION: ANTIFUNGAL AND HERBICIDAL COMPOUNDS THAT TARGET THE

TITLE OF INVENTION: FIRST COMMITTED STEP IN SPHINGOLIPID LONG-CHAIN

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: LOWE, PRICE, LEBLANC & BECKER

STREET: STE. 300, 99 CANAL CENTER PLAZA

CITY: Alexandria

STATE: Virginia

COUNTRY: USA

ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: DOS Text File

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/365,981

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/906,899

FILING DATE: 06/30/92

ATTORNEY/AGENT INFORMATION:

NAME: D.J. MILLS

REGISTRATION NUMBER: 34506

REFERENCE/DOCKET NUMBER: 434-028

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 684-1111

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 333

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2003, 22:56:01 ; Search time 58 seconds

(without alignments)
121.613 Million cell updates/sec

Title: SEQ1-T-AT-12347_COPY_12336_12358

Perfect score: 23
Sequence: 1 ggggcaatcttactctctcc 23

Scoring table:
Gapop 60.0, Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size: 12

Total number of hits satisfying chosen parameters: 122

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCITUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	60.9	35081	2	US-08-752-760A-1
2	13	56.5	100	1	US-07-795-859B-10
3	13	56.5	100	1	US-08-457-616-10
4	13	56.5	333	1	US-08-365-981-1
5	13	56.5	339	4	US-09-114-001C-96
6	13	56.5	339	4	US-09-114-001C-131
7	13	56.5	614	4	US-09-385-982-148
8	13	56.5	677	4	US-08-858-207A-167
9	13	56.5	752	4	US-09-175-658B-23
10	13	56.5	765	4	US-09-134-001C-1728
11	13	56.5	1299	3	US-08-793-035-3
12	13	56.5	2073	4	US-09-114-001C-1731
13	13	56.5	2418	4	US-08-549-515-2
14	13	56.5	3083	4	US-08-961-527-138
15	13	56.5	4051	3	US-08-549-515-1
16	13	56.5	4112	1	US-08-340-203A-2
17	13	56.5	4112	2	US-08-452-567-2
18	13	56.5	4112	2	US-08-452-427-2
19	13	56.5	4112	3	US-09-085-407-2
20	13	56.5	4616	1	US-08-340-203A-1
21	13	56.5	4616	2	US-08-452-567-1
22	13	56.5	4616	2	US-08-452-427-1
23	13	56.5	4616	3	US-09-085-407-1
24	13	56.5	10240	4	US-08-961-527-32
25	13	56.5	31728	4	US-09-453-702B-64
26	12	52.2	20	1	US-08-196-538-21
27	12	52.2	33	1	US-08-196-538-23

C 28	12	52.2	86	2	US-08-870-518-12	Sequence 12, Appl
C 29	12	52.2	255	2	US-08-870-518-29	Sequence 29, Appl
C 30	12	52.2	294	2	US-08-870-518-31	Sequence 31, Appl
C 31	12	52.2	296	4	US-09-370-838-86	Sequence 86, Appl
C 32	12	52.2	350	3	US-08-888-077A-32	Sequence 32, Appl
C 33	12	52.2	358	1	US-08-686-878A-3	Sequence 3, Appl
C 34	12	52.2	400	3	US-08-714-918-4	Sequence 4, Appl
C 35	12	52.2	400	4	US-09-265-315-4	Sequence 4, Appl
C 36	12	52.2	400	4	US-09-265-315-4	Sequence 4, Appl
C 37	12	52.2	400	4	US-09-266-417-4	Sequence 4, Appl
C 38	12	52.2	439	2	US-08-870-518-32	Sequence 32, Appl
C 39	12	52.2	505	4	US-08-991-789A-205	Sequence 205, App
C 40	12	52.2	505	4	US-09-062-451-205	Sequence 205, App
C 41	12	52.2	505	4	US-09-598-326-205	Sequence 205, App
C 42	12	52.2	515	4	US-09-586-719-29	Sequence 29, Appl
C 43	12	52.2	572	4	US-08-961-527-391	Sequence 391, App
C 44	12	52.2	611	4	US-09-328-111-375	Sequence 375, App
C 45	12	52.2	787	4	US-09-370-838-53	Sequence 53, Appl

ALIGNMENTS

RESULT 1
US-08-752-760A-1
Sequence 1, Application US/08752760A
Patent No. 5877011
GENERAL INFORMATION:
APPLICANT: Armentano, Donna
APPLICANT: Gregory, Richard J.
TITLE OF INVENTION: CHIMERIC ADENOVIRAL VECTORS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Baker & Bolls, L.L.P.
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,760A
FILING DATE: 20-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Seide, Rochelle K
REGISTRATION NUMBER: 32,300
REFERENCE/DOCKET NUMBER: A31385
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-705-5000
TELEFAX: 212-705-5020
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 35081 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: Linear
US-08-752-760A-1
Query Match 60.9% Score 14; DB 2; Length 35081;
Best Local Similarity 100.0%; Pred. No. 3; 6;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY 10 CTTACTTCGCTCC 23

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: January 11, 2003, 23:43:17 ; Search time 58 Seconds
(without alignments)
174.524 Million cell updates/sec

Title: SE01-T-AF-12347_COPY_12336_12358

Perfect score: 23

Sequence: 1 ggggcaatcttacttcgtcc 23

Scoring table: OLIGO_NUC

Gapop 60.0, Gapext 60.0

Searched:

389086 seqs, 220051671 residues

Word size: 12

Total number of hits satisfying chosen parameters: 242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: Published Applications_NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	15	65.2	462	10	US-09-815-343-62
C 2	15	65.2	630	9	US-10-149-256-3
C 3	14	60.9	660	10	US-09-974-300-3323
C 4	14	60.9	1695	9	US-09-938-842A-395
C 5	14	60.9	2000	9	US-09-938-842A-4537
C 6	14	60.9	2622	10	US-09-917-800A-1644
C 7	14	60.9	4780	10	US-09-886-241-1
C 8	14	60.9	17908	10	US-09-764-877-3619
C 9	14	60.9	35100	10	US-09-782-378A-26
C 10	13	56.5	197	10	US-09-983-965-5826
C 11	13	56.5	248	10	US-09-867-701-9
C 12	13	56.5	264	9	US-10-025-380-782
C 13	13	56.5	264	10	US-09-922-217-782
C 14	13	56.5	264	10	US-09-833-263-782
C 15	13	56.5	412	10	US-09-867-701-4222
C 16	13	56.5	423	9	US-09-796-692-7426
C 17	13	56.5	470	10	US-09-864-761-1432
C 18	13	56.5	474	10	US-09-864-761-14221
C 19	13	56.5	492	10	US-09-815-242-7159

C 20	13	56.5	526	10	US-09-778-320-130	Sequence 130, App
C 21	13	56.5	526	10	US-09-910-689-130	Sequence 130, App
C 22	13	56.5	526	12	US-10-010-742-130	Sequence 130, App
C 23	13	56.5	539	10	US-09-864-761-12502	Sequence 12502, A
C 24	13	56.5	546	9	US-10-046-935-205	Sequence 205, App
C 25	13	56.5	546	9	US-09-878-178-205	Sequence 205, App
C 26	13	56.5	600	9	US-09-796-692-8143	Sequence 8143, App
C 27	13	56.5	647	9	US-09-796-692-8521	Sequence 8521, App
C 28	13	56.5	803	10	US-09-800-729-60	Sequence 60, App
C 29	13	56.5	1370	10	US-09-887-576-536	Sequence 536, App
C 30	13	56.5	1374	10	US-09-887-576-525	Sequence 525, App
C 31	13	56.5	1542	12	US-10-044-090-698	Sequence 698, App
C 32	13	56.5	1628	10	US-09-799-777-124	Sequence 124, App
C 33	13	56.5	1821	10	US-09-800-729-59	Sequence 59, App
C 34	13	56.5	1873	10	US-09-820-893-15	Sequence 15, App
C 35	13	56.5	1875	9	US-09-938-842A-2959	Sequence 2959, App
C 36	13	56.5	1897	10	US-09-820-893-42	Sequence 42, App
C 37	13	56.5	1951	10	US-09-864-761-15729	Sequence 15729, A
C 38	13	56.5	1969	9	US-09-938-842A-5119	Sequence 5119, App
C 39	13	56.5	2000	9	US-09-938-842A-2962	Sequence 2962, App
C 40	13	56.5	2000	9	US-09-938-842A-5263	Sequence 5263, App
C 41	13	56.5	2003	10	US-09-887-576-7	Sequence 7, App
C 42	13	56.5	2025	10	US-09-925-299-117	Sequence 117, App
C 43	13	56.5	2270	10	US-09-822-830A-45	Sequence 45, App
C 44	13	56.5	2288	10	US-09-800-729-24	Sequence 24, App
C 45	13	56.5	2323	12	US-10-044-090-530	Sequence 530, App

ALIGNMENTS

RESULT 1
US-09-815-343-62/C
Sequence 62, Application US/09815343
Patent No. US2001005596A1
GENERAL INFORMATION:
APPLICANT: Meagher, Madeleine
APPLICANT: Xu, Jiangchun
APPLICANT: King, Gordon E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
FILE REFERENCE: 210121.504
CURRENT APPLICATION NUMBER: US/09/815,343
CURRENT FILING DATE: 2001-03-22
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 62
LENGTH: 462
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(462)
OTHER INFORMATION: n = A,T,C or G
US-09-815-343-62

Query Match 65.2%; Score 15; DB 10; Length 462;
Best local Similarity 100.0%; Pred. No. 1.7;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGGCAATCTTACTT 16
DB 313 GGGCAATCTTACTT 299

RESULT 2
US-10-149-256-3/C
Sequence 3, Application US/10149256
Publication No. US20030003488A1
GENERAL INFORMATION:
APPLICANT: Ramakrishnan, Shyam
TITLE OF INVENTION: Regulation of Human Mitochondrial
TITLE OF INVENTION: Deformylase

FILE REFERENCE: 004974.00739
CURRENT APPLICATION NUMBER: US/10/149,256
CURRENT FILING DATE: 2002-06-10
PRIOR APPLICATION NUMBER: PCT/EP00/12110
PRIOR FILING DATE: 2000-12-01
PRIOR APPLICATION NUMBER: US 60/239,106
PRIOR FILING DATE: 2000-10-11
PRIOR APPLICATION NUMBER: US 60/169,615
PRIOR FILING DATE: 1999-12-08
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 630
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(630)
OTHER INFORMATION: n = A,T,C or G
US-10-149-256-3

Query Match 65.2%; Score 15; DB 9; Length 630;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGGCAATCTTACTT 16
DB 338 GGGCAATCTTACTT 324

RESULT 3
US-09-974-300-3323
Sequence 3323, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berkta, Randy M.
APPLICANT: Clausen, Ib Groth
TITLE OF INVENTION: Methods for Monitoring Multiple Gene
FILE REFERENCE: 10085.500-US
CURRENT APPLICATION NUMBER: US/09/974,300
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3323
LENGTH: 660
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-3323

Query Match 60.9%; Score 14; DB 10; Length 660;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AAATCTTACTTTCG 19
DB 371 AAATCTTACTTTCG 384

RESULT 4
US-09-938-842A-395
Sequence 395, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

TITLE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 395
LENGTH: 1695
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-395

Query Match 60.9%; Score 14; DB 9; Length 1695;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AAATCTTACTTTCG 19
DB 330 AAATCTTACTTTCG 343

RESULT 5
US-09-938-842A-4537
Sequence 4537, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 4537
LENGTH: 2000
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-4537

Query Match 60.9%; Score 14; DB 9; Length 2000;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CAATCTTACTTTCG 18
DB 1085 CAATCTTACTTTCG 1098

RESULT 6
US-09-917-800A-1644/C
Sequence 1644, Application US/09917800A
Patent No. US20020119462A1
GENERAL INFORMATION:
APPLICANT: Mendrick, Donna
APPLICANT: Porter, Mark
APPLICANT: Johnson, Kory
APPLICANT: Castle, Arthur
APPLICANT: Elashoff, Michael
APPLICANT: Gene Logic, Inc.

```

: TITLE OF INVENTION: Molecular Toxicology Modeling
: FILE REFERENCE: 44921-5038-US
: CURRENT APPLICATION NUMBER: US/09/917,800A
: CURRENT FILING DATE: 2001-07-31
: PRIOR APPLICATION NUMBER: US 60/222,040
: PRIOR FILING DATE: 2000-07-31
: PRIOR APPLICATION NUMBER: US 60/222,880
: PRIOR FILING DATE: 2000-11-02
: PRIOR APPLICATION NUMBER: US 60/290,029
: PRIOR FILING DATE: 2001-05-11
: PRIOR APPLICATION NUMBER: US 60/290,645
: PRIOR FILING DATE: 2001-05-15
: PRIOR APPLICATION NUMBER: US 60/292,336
: PRIOR FILING DATE: 2001-05-22
: PRIOR APPLICATION NUMBER: US 60/295,798
: PRIOR FILING DATE: 2001-06-06
: PRIOR APPLICATION NUMBER: US 60/297,457
: PRIOR FILING DATE: 2001-06-13
: PRIOR APPLICATION NUMBER: US 60/298,884
: PRIOR FILING DATE: 2001-06-19
: PRIOR APPLICATION NUMBER: US 60/303,459
: PRIOR FILING DATE: 2001-07-09
: NUMBER OF SEQ ID NOS: 1740
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1644
: LENGTH: 2622
: TYPE: DNA
: ORGANISM: Rattus norvegicus
: FEATURE:
: OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_017208
US-09-917-800A-1644
```

```

Query Match      60.9%; Score 14; DB 10; Length 2622;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CAAATCTTACTTTC 18
    |||||
DB 820 CAAATCTTACTTTC 807
```

```

RESULT 7
US-09-886-241-1
: Sequence 1, Application US/09886241
: Patent No. US20020068691A1
: GENERAL INFORMATION:
: APPLICANT: Salceda, Susana
: APPLICANT: Catterkey, Robert
: TITLE OF INVENTION: Method of Diagnosing, Monitoring, Staging, Imaging and
: FILE REFERENCE: DEX-0209
: CURRENT APPLICATION NUMBER: US/09/886,241
: CURRENT FILING DATE: 2001-06-21
: PRIOR APPLICATION NUMBER: 60/213,084
: PRIOR FILING DATE: 2000-06-21
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1
: LENGTH: 4780
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-886-241-1
```

```

Query Match      60.9%; Score 14; DB 10; Length 4780;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 3 GCGAATCTTACTT 16
    |||||
DB 2091 GCGAATCTTACTT 2104
```

RESULT 8

```

US-09-764-877-3619/c
: Sequence 3619, Application US/09764877
: Patent No. US20020147140A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PC005
: CURRENT APPLICATION NUMBER: US/09/764,877
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - refer to PALM or file wrapper
: NUMBER OF SEQ ID NOS: 4031
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 3619
: LENGTH: 17908
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (51)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (69)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (85)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (97)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (98)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (110)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (123)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (149)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (400)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (555)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (686)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (718)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (888)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (1289)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (2284)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (2354)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (2456)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (2641)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (2669)
```

```

OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (3176)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (3186)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (3203)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (3217)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (3234)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (3412)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (3431)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4238)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4342)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4351)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4360)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4434)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4435)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4445)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4459)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4470)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4472)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4475)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4486)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4494)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4503)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4507)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4515)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4524)
OTHER INFORMATION: n equals a,t,g, or c

```

```

NAME/KEY: SITE
LOCATION: (4538)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4558)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4616)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4705)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (4754)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (5298)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (5311)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (5513)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (6363)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (7252)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (8375)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (8540)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (8878)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (9805)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (9839)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (10029)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (10046)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (10389)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE

```

Query Match 60.9%; Score 14; DB 10; Length 17908;
 Best Local Similarity 100.0%; Pred. No. 7.3;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 5 CAATCTTACTTTC 18
 Db 17586 CAATCTTACTTTC 17573

RESULT 9
 US-09-782-378A-26
 Sequence 26, Application US/09782378A
 Patent No. US20020102731A1
 GENERAL INFORMATION:
 APPLICANT: Hearing, Patrick
 APPLICANT: Bahou, Wadie
 APPLICANT: Sandation, Ziv

```

; APPLICANT: Gatenko, Dmitri
; TITLE OF INVENTION: Adenoviral Vectors
; FILE REFERENCE: STONYB-04970
; CURRENT APPLICATION NUMBER: US/09/782,378A
; CURRENT FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/237,747
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 26
; LENGTH: 35100
; TYPE: DNA
; ORGANISM: Human adenovirus type 17
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (25184)..(25184)
; OTHER INFORMATION: The n at this position can be a, c, t, or g.
US-09-782-378A-26

```

```

Query Match          60.9%; Score 14; DB 10; Length 35100;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 10 CTTACTTCCGCTCC 23
Db 21170 CTTACTTCCGCTCC 21183

```

```

RESULT 10
US-09-983-965-5826/c
; Sequence 5826, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nadappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 5826
; LENGTH: 197
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 60-LIB34-059-Q1-F1-G12
US-09-983-965-5826

```

```

Query Match          56.5%; Score 13; DB 10; Length 197;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 5 CAATCTTACTTT 17
Db 145 CAATCTTACTTT 133

```

```

RESULT 11
US-09-867-701-9
; Sequence 9, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

```

```

; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 248
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9

```

```

Query Match          56.5%; Score 14; DB 10; Length 248;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 6 AATCTTACTTTC 18
Db 191 AATCTTACTTTC 203

```

```

RESULT 12
US-10-025-480-782
; Sequence 782, Application US/10025480
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeline Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yiqiu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skolky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C14
; CURRENT APPLICATION NUMBER: US/10/025,480
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 782
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 47, 69, 195, 262
; OTHER INFORMATION: n = A,T,C or G
US-10-025-480-782

```

```

Query Match          56.5%; Score 13; DB 9; Length 264;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GGGGCAATCTTA 13
Db 172 GGGGCAATCTTA 184

```

```

RESULT 13
US-09-922-217-782
; Sequence 782, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secretist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 782
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 47, 69, 195, 262
; OTHER INFORMATION: n = A,T,C or G
US-09-922-217-782

```

```

Query Match          56.5%; Score 13; DB 10; Length 264;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GGGGCAATCTTA 13
   |||
DB 172 GGGGCAATCTTA 184

```

```

RESULT 14
US-09-833-263-782
; Sequence 782, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.
; APPLICANT: Meagher, Madeleine J.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833,263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 782
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(264)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-263-782

```

```

Query Match          56.5%; Score 13; DB 10; Length 264;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GGGGCAATCTTA 13
   |||
DB 172 GGGGCAATCTTA 184

```

```

RESULT 15
US-09-867-701-4322/c
; Sequence 4322, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4322
; LENGTH: 412
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-4322

```

```

Query Match          56.5%; Score 13; DB 10; Length 412;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 3 GGCAATCTTACT 15
   |||
DB 73 GGCAATCTTACT 61

```

```

Search completed: January 12, 2003, 00:28:41
Job time : 73 secs

```